

Aplikace prvků umělé inteligence při hodnocení polohových měření laser-interferometrem

Bc. Filip Kamák

Diplomová práce
2025



Univerzita Tomáše Bati ve Zlíně
Fakulta technologická

Univerzita Tomáše Bati ve Zlíně

Fakulta technologická

Ústav výrobního inženýrství

Akademický rok: 2024/2025

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	Bc. Filip Kamák
Osobní číslo:	T22364
Studijní program:	N0788A270007 Řízení jakosti
Forma studia:	Prezenční
Téma práce:	Aplikace prvků umělé inteligence při hodnocení polohových měření laser-interferometrem

Zásady pro vypracování

- Vypracování rešerše na zadané téma
- Proměření polohových dat laser-interferometrem
- Nalezení vhodné struktury Neuronové sítě, včetně její optimalizace

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam doporučené literatury:

MELOUN, Milan a Jiří MILITKÝ. *Kompendium statistického zpracování dat*. Praha: Karolinum, 2012. ISBN 978-80-246-2196-8.
WHITEHOUSE, D. J. *Handbook of surface and nanometrology*. 2nd ed. Boca Raton: CRC Press, c2011. ISBN 978-1-4200-8201-2.
AGGARWAL, Charu C. *Neural networks and deep learning: a textbook*. Cham: Springer, [2018].
GRAUPE, Daniel. *Deep learning neural networks: design and case studies*. Singapore: World scientific, [2016]. ISBN 978-981-3146-44-0.

Vedoucí diplomové práce: **prof. Dr. Ing. Vladimír Pata**
Ústav výrobního inženýrství

Datum zadání diplomové práce: **2. ledna 2025**
Termín odevzdání diplomové práce: **9. května 2025**

prof. Ing. Roman Čermák, Ph.D. v.r.
děkan

L.S.

doc. Ing. Martin Bednařík, Ph.D. v.r.
ředitel ústavu

Ve Zlíně dne 10. února 2025

PROHLÁŠENÍ AUTORA DIPLOMOVÉ PRÁCE

Beru na vědomí, že

- odevzdáním diplomové práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb., v platném znění bez ohledu na výsledek obhajoby;
- diplomová práce bude uložena v elektronické podobě v univerzitním informačním systému a bude dostupná k nahlédnutí;
- jedno vyhotovení diplomové práce v listinné podobě bude ponecháno Univerzitě Tomáše Bati ve Zlíně k uložení;
- na moji diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- podle § 60 odst. 1 autorského zákona má Univerzita Tomáše Bati ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- podle § 60 odst. 2 a 3 mohu užít své dílo – diplomovou práci – nebo poskytnout licenci k jejímu využití jen s předchozím písemným souhlasem Univerzity Tomáše Bati ve Zlíně, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše);
- pokud bylo k vypracování diplomové práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tj. k nekomerčnímu využití), nelze výsledky diplomové práce využít ke komerčním účelům;
- pokud je výstupem diplomové práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá; neodevzdání této součásti může být důvodem k neobhájení práce.

Prohlašuji, že

- jsem na diplomové práci pracoval(a) samostatně a použitou literaturu jsem řádně citoval(a); v případě publikace výsledků budu uveden(a) jako spoluautor;
- odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou obsahově totožné.

Ve Zlíně, dne

.....

podpis autora

ABSTRAKT

Tato diplomová práce se zabývá vyhodnocením naměřených dat pomocí regresní analýzy, přičemž jsou pro vyhodnocení využívány prvky umělé inteligence. V teoretické části této práce jsou popsány základní principy statistického vyhodnocování dat, fungování neuronových sítí, jazykový model ChatGPT a problematika měření robotických paží.

Praktická část této práce je zaměřena na vyhodnocení naměřených dat a následný trénink jazykového modelu ChatGPT, který následně poslouží i běžným uživatelům bez jakýchkoliv znalostí statistiky jako pomocník při provádění regresních analýz.

Klíčová slova: regresní analýza, umělá inteligence, statistické zpracování dat, jazykový model ChatGPT

ABSTRACT

This thesis deals with the evaluation of measured data by means of regression analysis using artificial intelligence elements. The theoretical part of this thesis describes the basic principles of statistical data evaluation, the operation of neural networks, the ChatGPT language model and the problem of measuring robotic arms.

The practical part of this thesis is focused on the evaluation of measured data and subsequent training of the ChatGPT language model, which will then serve as an aid to ordinary users without any knowledge of statistics in performing regression analyses.

Keywords: regression analysis, artificial intelligence, statistical data processing, ChatGPT language model

V této části diplomové práce bych rád poděkoval své rodině, která mě po celou dobu vysokoškolského studia podporovala. Zvláštní poděkování patří mým prarodičům, kteří stáli vždy při mně a byli mi velkou oporou. Bez jejich pomoci a povzbuzení, stejně jako bez podpory celé mé rodiny, by nebylo možné těchto milníků dosáhnout.

Velké díky patří také vedoucímu mé diplomové práce, panu prof. Dr. Ing. Vladimíru Patovi, jehož odborné znalosti a rady byly klíčové pro úspěšné zvládnutí celého studia.

OBSAH

ÚVOD.....	13
I TEORETICKÁ ČÁST	14
1 MĚŘENÍ VÝROBNÍCH STROJŮ	15
1.1 MĚŘENÍ JAKO POJEM	15
1.1.1 Fundamentální metrologie	15
1.1.2 Technická metrologie	15
1.1.3 Legální metrologie	15
1.1.4 Všeobecná metrologie	16
1.2 MĚŘÍCÍ PRINCIPY A POSTUPY	16
1.3 GEOMETRICKÁ PŘESNOST PŘI POLOHOVÁNÍ	16
1.3.1 Kinematické chyby	16
1.3.2 Mechanické opotřebení	17
2 LASERY	18
2.1 DRUHY LASERŮ	18
2.1.1 Lasery plynové HeNe	18
2.2 INTERFERENCE SVĚTLA	19
2.3 INTERFEROMETRIE	19
2.4 LASEROVÝ INTERFEROMETR	19
2.4.1 Michelsonův interferometr	20
3 NEURONOVÉ SÍTĚ	21
3.1 ARCHITEKTURA NEURONOVÝCH SÍTÍ	21
3.1.1 Neuron	21
3.1.2 Vrstvy neuronové sítě	22
3.1.3 Feedforward neural network	22
3.1.4 Recurrent neural network	23
3.1.5 Convolutional neural network	23
3.2 UČENÍ NEURONOVÝCH SÍTÍ	23
3.2.1 Princip backpropagation	24
4 CHATGPT	25
4.1 SPOLEČNOST OPENAI	25
4.2 DRUHY PŘEDPLATNÉHO	25
4.2.1 Free verze	26
4.2.2 Předplatné Plus	26
4.2.3 Předplatné Team	27
4.2.4 Verze Pro	28
4.2.5 Předplatné Enterprise	28
4.3 VERZE JAZYKOVÝCH MODELŮ CHATGPT	29

4.3.1	GPT 3.5	29
4.3.2	GPT-4o	29
4.3.3	GPT-4o mini.....	29
4.3.4	Model o1 a o1-mini	29
5	HODNOCENÍ DAT NA STATISTICKÉ BÁZI.....	31
5.1	METODIKA EDA	31
5.1.1	Určení typu dat	31
5.1.2	Grafické zobrazení hodnot	32
5.1.3	Metody vyjádření a interpretace dat.....	32
5.2	REGRESNÍ ANALÝZA	33
5.2.1	Pravidla pro sestavení lineární regrese.....	33
6	HODNOCENÍ DAT NA STATISTICKÉ BÁZI.....	35
6.1	ZDROJE VZNIKU A TYPY CHYB PŘI MĚŘENÍ	35
6.1.1	Chyby při měření.....	35
6.1.2	Chyba náhodná, systematická a hrubá	35
6.2	VLIV CHYB NA MĚŘENÍ POMOCÍ LASEROVÝCH INTERFEROMETRŮ.....	36
6.2.1	Změna teploty.....	36
6.2.2	Vibrace	36
6.2.3	Optické chyby a aberace.....	36
6.2.4	Interferenční efekty a chybové signály	36
6.3	KALIBRACE ROBOTŮ.....	36
6.3.1	Proces provádění kalibrace.....	37
7	ROBOTIKA	38
7.1	PRŮMYSLOVÝ ROBOT IRB 14000 - YuMi	38
7.2	POPIS ROBOTA YUMI Z HLEDISKA JEHO KALIBRACE	39
7.2.1	Kalibrace při změně hodnot resolveru.....	39
7.2.2	Kalibrace při ztrátě paměti otáčkoměru	39
7.2.3	Při přestavbě robota.....	40
7.3	KALIBRACE ROBOTICKÝCH PAŽÍ	40
7.3.1	CalibWare.....	40
7.3.2	YuMi AbsAcc Recovery	40
II	PRAKTICKÁ ČÁST	42
8	CÍLE PRÁCE.....	43
9	EXPERIMENT	44
9.1	MĚŘENÍ DAT	44
9.2	ZPRACOVÁNÍ NAMĚŘENÝCH DAT	44
9.2.1	Explorační analýza dat – EDA	45
9.2.2	Regresní analýza.....	46
9.3	VYTVOŘENÍ PROMPTU POMOCÍ CHATGPT.....	49

9.3.1	První experiment	49
9.3.2	Druhý experiment – první skutečný prompt	51
9.3.3	Třetí experiment – rozdělení celého procesu	55
9.3.4	Čtvrtý experiment – ošetření vstupů a ladění textů.....	65
9.3.5	První chyba během analýzy dat.....	79
9.3.6	Finalizace promptu.....	80
9.3.7	Výsledný prompt	101
9.3.8	Použití v bezplatné verzi ChatGPT	103
ZÁVĚR		104
SEZNAM POUŽITÉ LITERATURY.....		105
SEZNAM OBRÁZKŮ		109
SEZNAM TABULEK.....		110
SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK.....		111

ÚVOD

Technologický pokrok a automatizace v mnohých průmyslových odvětvích mají za následek stále vyšší nároky na zpracování a vyhodnocení dat. Modernizace výrobních procesů způsobila, že výrobní systémy generují obrovské množství dat a informací, jejichž správná interpretace bývá zásadní pro zajištění přesnosti a optimalizace výrobních procesů. Ve výrobním průmyslu hraje klíčovou roli nejen výroba, ale i měření a analýza jednotlivých procesů, díky které je možné včasné identifikovat odchylky či chyby.

S příchodem prvních neuronových sítí se dynamicky rozvíjejí technologie umělé inteligence, které se stávají každodenními pomocníky při řešení komplexních úloh a problémů. Velmi revoluční se v této oblasti stala společnost OpenAI s jejím jazykovým modelem ChatGPT, který představuje zcela novou dimenzi při řešení jakéhokoliv problému včetně využití pro statistickou analýzu dat. Tento jazykový model umožňuje uživatelům provádět nejrůznější datové analýzy včetně vyhodnocení naměřených dat a jejich zpracování.

Diplomová práce má za cíl zpracovat naměřená data z hlediska regresní analýzy a na základě těchto poznatků vycvičit jazykový model společnosti OpenAI – ChatGPT 4o tak, aby data zpracoval a vyhodnotil stejným způsobem. Výsledkem celé práce bude optimalizovaný prompt – program, díky kterému bude možné provést regresní analýzu bez hlubších znalostí statistiky. Hlavní výhodou této metody je rychlost, efektivita a úspora peněz, protože uživatel k vyhodnocení potřebuje pouze internet a předplacenou verzi ChatGPT 4o.

I TEORETICKÁ ČÁST

1 MĚŘENÍ VÝROBNÍCH STROJŮ

Měření je nezbytnou součástí mnoha různých odvětví od průmyslu až po vědecký výzkum, protože umožňuje porovnávání a kontrolu kvality. Měření výrobních strojů hraje významnou roli v průmyslu především kvůli požadavkům, jako přesnost, opakovatelnost a efektivita, které jsou na výrobní stroje kladeny. [1], [2], [3]

1.1 Měření jako pojem

Věda zabývající se měřením se nazývá metrologie. Měření jako takové je zkoumání a porovnávání naměřených číselných hodnot s obecně přijatelnou hodnotou, ovšem nemusí se jednat výhradně o porovnávání naměřených hodnot. Do metrologie spadají i související procesy, kterými může být kalibrace měřících přístrojů, standardizace a určování nejistot měření. Podle povahy a účelu měřené veličiny je možné metrologii rozdělit do následujících odvětví.

1.1.1 Fundamentální metrologie

Fundamentální metrologie stanovuje základní principy měření, soustavu měřících jednotek, určuje etalony a soustavy fyzikálních konstant. Toto odvětví metrologie je zodpovědné za udržování mezinárodních standardů měření.

1.1.2 Technická metrologie

V některých případech je technická metrologie označovaná i jako průmyslová metrologie jejímž úkolem je praktická aplikace měřících technik ve výrobě. Především se jedná o kalibraci přístrojů a sledování jejich přesnosti. V průmyslu má obrovský význam, protože zajišťuje, že výroba probíhá ve stanovených tolerancích, čím napomáhá ke zvýšení efektivit a snížení zmetkovitosti.

1.1.3 Legální metrologie

Pod legální metrologii spadá měření v oblasti obchodu, bezpečnosti a ochrany spotřebitele. Cílem tohoto odvětví metrologie je zajistit, aby všechna měřící zařízení splňovala požadavky na přesnost a nedocházelo k neshodám, které by mohly vést ke snížení důvěry spotřebitelů. Jednou z oblastí může být např. vážení produktů v maloobchodě, nebo měření objemu natankovaného paliva na čerpacích stanicích.

1.1.4 Všeobecná metrologie

Všeobecná metrologie zahrnuje zbylé měřicí aktivity, které přímo nespadají do výše zmíněných oblastí metrologie. Zabývá se především obecným principem měření, nejistotami měření, metodologií metrologických procesů a jinými principy spojenými s metrologií. [1], [2]

1.2 Měřicí principy a postupy

Měření výrobních strojů je jedním ze zásadních aspektů v oblasti průmyslu, protože má vliv na přesnost, opakovatelnost a optimální výkon výrobního prostředí. K změření přesnosti výrobních strojů je možné využít různé metody, které lze rozdělit na přímé a nepřímé.

Mezi přímé měřicí metody lze zahrnout takové techniky, které měří konkrétní fyzikální veličiny, jako jsou různé délky nebo úhly. U těchto metod není nutné provádět další převod či transformace výsledku. Do přímých měřících metod lze zařadit měření pomocí laser interferometru a optických snímačů. Obrovskou výhodou uvedených technik je vysoká přesnost. [4], [5]

U nepřímých metod měření se výsledky odvozují od sekundárních parametrů. Pokud bychom chtěli vyhodnotit odchylku polohy robota, je možné využít snímače síly nebo tenzometry, které porovnají skutečné odchylky od předem definovaných trajektorií. Nevýhodou nepřímých metod je nutnost zpracování dat a převod údajů ze senzorů do hodnot polohy. [5], [6]

1.3 Geometrická přesnost při polohování

Geometrická přesnost při polohování zajišťuje přesně definovaný pohyb stroje či nástroje v předem definovaném souřadnicovém systému. V dnešním průmyslu hraje přesnost při polohování zásadní roli pro zajištění kvality a konzistence výrobních procesů. Mezi hlavní faktory ovlivňující geometrickou přesnost lze zařadit kinematické chyby, montážní odchylky, mechanické opotřebení a jiné vnější vlivy. [3]

1.3.1 Kinematické chyby

Kinematické chyby jsou spojené s pohybem robota a vyskytují se jako následek nedokonalostí v jeho konstrukci či montáži. Tyto chyby způsobují odchylky mezi skutečně uraženou a teoretickou vzdáleností, což může vést k chybnému polohování, jehož následky mohou negativně ovlivnit výrobu. [3], [4]

1.3.2 Mechanické opotřebení

Mechanické opotřebení je spojeno s nepřetržitým provozem výrobních strojů. Postupem času dochází u strojů vlivem opakovatelnosti k degradaci mechanických součástí a následnému vzniku vůlí, které mohou negativně ovlivnit výrobu. [4], [5]

2 LASERY

Lasery jsou zařízení, která přeměňují elektrickou, chemickou, nebo nekoherentní optickou energii na koherentní optické záření. Ke svému fungování využívají zesílení světla pomocí stimulované emise záření. Koherence laserového záření může být prostorová i spektrální, což znamená, že je světlo uspořádáno směrově i frekvenčně. Prostorová koherence ovlivňuje především směr světelného paprsku a spektrální koherence má za následek čistotu barvy záření. Každé laserové zařízení obsahuje atomární nebo molekulární zesilovací médium, které je umístěno v optickém rezonátoru či dutině. Princip fungování spočívá v oscilaci atomů nebo molekul v zesilovacím médiu na optických frekvencích vlivem elektrické nebo optické energie. K oscilaci dochází za extrémně vysokých frekvencích. Vysoce výkonné lasery s úzkou čárovou šířkou dosahují šířky spektrální čáry přibližně kolem 300 MHz. Nízkovýkonné lasery s úzkou čárovou šířkou mohou dosahovat přibližně 100 kHz a u stabilizovaných laserů se jedná o 1 kHz a méně. Lasery s úzkou čárovou šířkou se také označují jako laserové oscilátory a nacházejí využití zejména ve spektroskopii nebo metrologii. [7] [8]

2.1 Druhy laserů

V dnešní době existuje velké množství různých druhů laserových zařízení. Každé laserové zařízení funguje na trochu jiném principu oproti ostatním zařízením a z tohoto důvodu lze lasery rozřadit podle různých charakteristik. Nejčastějším kritériem, podle kterého lze lasery rozdělit, je aktivní prostředí. Dle aktivního prostředí je možné rozlišit lasery plynové, barvivové (dye lasery), pevnolátkové, polovodičové a další. [7]

2.1.1 Lasery plynové HeNe

Helium-neonové lasery jsou jedním z nejznámějších druhů plynových laserů. Aktivní prostředí těchto laserů je tvořeno plynou směsí hélia a neonu s tlakem okolo 400 Pa. K elektrickému výboji dochází v trubici, která má dvě elektrody. Princip fungování je založen na excitaci atomů hélia, u kterých dochází k přenosu excitační energie na atomy neonu, čímž dochází k laserovému přechodu. Výkon plynových laserů HeNe se pohybuje v řádu několika desetin až stovek miliwattů a pracuje především kontinuálně. [8]

2.2 Interference světla

Interference světla je fyzikální jev, při kterém dochází ke střetnutí, nebo prolnutí dvou a více světelných vln, což následně vede ke vzniku konstruktivní nebo destruktivní interference. Rozhodnutí, zda se ve výsledku bude jednat o konstruktivní nebo destruktivní interferenci závisí na jejich fázovém rozdílu. V důsledku odrazů vzniká při použití monochromatického světla interferenční signál. Vyslaný signál má oproti odraženému signálu vyšší posun a nižší kontrast. Hlavní výhodou užití bílého světla je jeho krátká koherenční délka. Takto krátká koherenční délka má za následek, že k interferenci dochází pouze při malém optickém rozdílu drah. Tento fyzikální jev nachází uplatnění především v optických přístrojích, spektroskopii, metrologii a holografii. [9], [10], [11]

2.3 Interferometrie

Interferometrie je měřicí metoda, která nachází uplatnění v mnoha oblastech např. v automobilovém, leteckém nebo lodním průmyslu. Výhodou bezkontaktní měřicí techniky oproti jiným technikám je zachování původních vzorků, protože během měření nedochází k jejich poškození. Princip interferometrie je založen na interferenci světla, které slouží k přesnému určení vzdáleností pomocí změn v optické dráze paprsků. Původní paprsek je rozdělen na dva svazky s rozdílnými optickými drahami a ve výsledku je opět spojen. Výsledná interference vytváří interferenční obrazec, díky kterému je možné určit přesné vzdálenosti s přesností na nanometry. Tohoto principu využívá i Michelsonův interferometr, který je více rozebrán v následující kapitole. [7], [9]

2.4 Laserový interferometr

Interferometr je zařízení využívající principu interference světla k přesnému měření s přesností nanometrů. Jak bylo popsáno výše, funguje na principu interferometrie, kdy dochází k rozdělení jednoho paprsku do dvou a následnému porovnání jejich optických drah. Každý interferometr je složen z několika zásadních komponentů, které zajišťují jeho funkci.

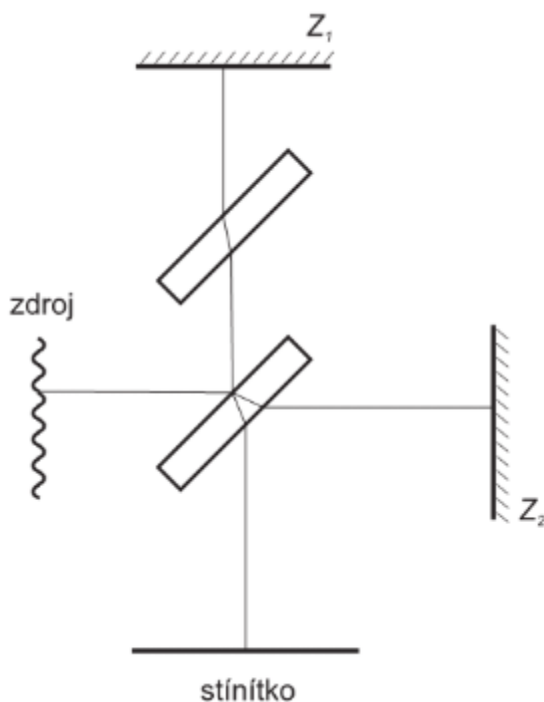
- Zdroj koherentního záření – zdrojem koherentního záření je nejčastěji laserová hlava, která vytváří stabilní světelný paprsek.
- Polopropustné zrcadlo slouží k rozdělení paprsku z laserové hlavy do dvou – měřicího a referenčního. Měřicí paprsek je směřován na měřený objekt, odkud

se za pomoci lineárního odražeče odráží zpět do detektoru a referenční paprsek zůstává uvnitř systému, přičemž je jeho optická dráha konstantní.

- Detektor – v detektoru dochází k opětovnému spojení obou paprsků a vyhodnocení rozdílů mezi nimi. Následnou analýzou interferenčního obrazce je možné vyhodnotit rozdíly v optických drahách. [9]

2.4.1 Michelsonův interferometr

Michelsonův interferometr je optické zařízení na bázi LCI – nízko-koherentní spektrometrie schopné v disperzním prostředí pomocí fázového sklonu měřit na úrovni nanometrů. Princip jeho fungování je založen na interferenci světla. Konkrétně u Michelsonova interferometru dochází k rozdělení jednoho světelného paprsku, který vychází z laserové hlavy pomocí polopropustného zrcadla do dvou (měřicího a referenčního). Měřicí paprsek je zaměřen na měřený objekt a odráží se zpět do detektoru, zatímco referenční paprsek slouží pro vzájemné porovnání optických drah. Rozdíl ve změně délky optické dráhy je možné zpozorovat pomocí interferenčního obrazce. Michelsonův interferometr v dnešní době nachází především díky své přesnosti uplatnění v mnoha odvětvích od využití v automobilovém průmyslu až po vesmírný průmysl a vývoj. Jedním z příkladů může být využití na palubě družice ICON, kde slouží k měření termosférických větrů. [7], [12], [13]



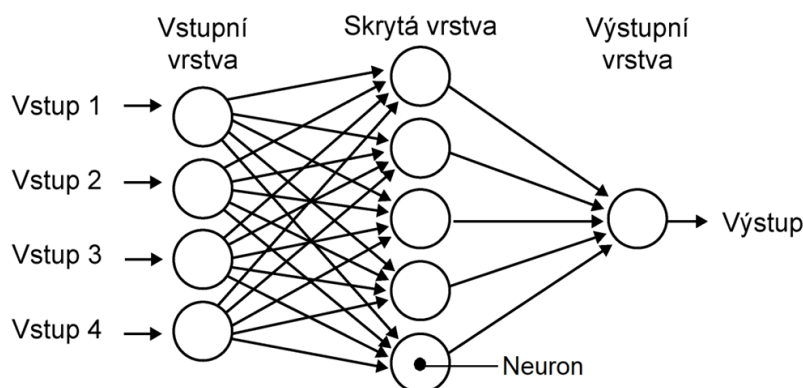
Obrázek 1. Princip fungování Michelsonova interferometru [8]

3 NEURONOVÉ SÍTĚ

Neuronové sítě představují zásadní pokrok v oblasti umělé inteligence. Fungují na podobném principu, jako lidská mysl. Díky svým jedinečným výpočetním vlastnostem nachází využití téměř ve všech odvětvích od výrobního průmyslu až po usnadnění administrativní činnosti. Podobně jako biologické neurony dokážou rozpoznat vzorce a vztahy mezi jednotlivými daty, které by byly tradičními metodami nerozpoznatelné. Jejich univerzální využití a neustálý vývoj může mít v budoucnu obrovský dopad na technologie a výzkum. [14], [15], [16]

3.1 Architektura neuronových sítí

Neuronové sítě jsou výpočetní modely inspirované procesy, odehrávajícími se v lidském mozku. Skládají se ze vzájemně propojených prvků – neuronů. Cílem neuronových sítí je identifikovat vzájemné spojitosti v datech, které umožňují např. rozpoznávat prvky nacházející se na obrázcích a vytvářet prediktivní analýzy. [14], [16], [17]



Obrázek 2. Architektura neuronových sítí

3.1.1 Neuron

Neuron – někdy nazýván jako uzel, je základní stavební jednotkou každé sítě a jeho činnost spočívá v identifikování vzorců a vztahů v datech pomocí matematických funkcí. Každý neuron v síti zpracovává vstupní data, kterým následně přiřadí různé váhy a zkreslení, čímž vzniká výstup. Neurony jsou vzájemně seřazeny do vrstev, v nichž probíhá zpracování vstupních dat a vzniká výstup, který je přenesen do další vrstvy. [17], [18], [19]

3.1.2 Vrstvy neuronové sítě

Podle uspořádání a pozice, kde se vrstva nachází, je možné rozlišit:

- vstupní vrstvu;
- skrytou vrstvu;
- výstupní vrstvu.

Vstupní (první) vrstva přijímá nezpracovaná data z vnějšího prostředí. Každý uzel nebo neuron v této vrstvě reprezentuje vlastnost vstupních dat, kterými mohou být určité hodnoty či intenzity pixelů v případě obrázků. První vrstva přenese tyto vlastnosti vstupních dat do další vrstvy bez provádění jakýchkoliv výpočtů a úprav.

Ve skrytých vrstvách dochází k výpočtům pomocí matematických funkcí, zkreslení a přiřazení různých vah. Tento výstup následně přechází do další vrstvy, která provádí totožné operace. Čím více skrytých vrstev neuronová síť obsahuje, tím lépe dokáže řešit složitější problémy. Celý tento koncept lze nastínit na následujícím příkladu s neuronovou sítí sloužící k rozpoznávání jednotlivých prvků na obrázku tzv. CNN – „Konvoluční neuronové sítě“. Počáteční skryté vrstvy umožňují síti detekovat jednoduché vzory, kterými mohou být přechody mezi světlými a tmavými oblastmi a hranami objektů. Oproti tomu hlubší vrstvy už dokážou rozpoznat složité vzory.

Úkolem výstupní vrstvy neuronové sítě je transformovat naučené vlastnosti do konečné předpovědi. V případě klasifikačních úloh odpovídá každý neuron pouze jedné výstupní třídě, přičemž model pro každou třídu vypisuje hodnoty, které určují konečnou předpověď. Počet neuronů je shodný s počtem cílových tříd. U regresních úloh je v mnoha případech ve výstupní vrstvě obsažen pouze jeden neuron, který předpovídá relevantní spojitou hodnotu pro daný problém. [14], [17], [18], [20]

3.1.3 Feedforward neural network

Feedforward neural network (FNN) v překladu dopředná neuronová síť je jednou ze základních typů neuronových sítí a využívá se k vytvoření a naučení pokročilejších sítí. Tato síť převádí data, která mohou proudit pouze jedním směrem, ze vstupu na výstup bez zacyklení či smyček. I když se jedná o jednu z nejjednodušších struktur, může obsahovat velké množství skrytých vrstev a spojení mezi těmito vrstvami může být složité. Dopřednou neuronovou síť lze využít pro různé úlohy, kterými mohou být regresní analýzy, rozpoznávání obrázků nebo klasifikace. [17], [21]

3.1.4 Recurrent neural network

Recurrent neural network (RNN) v českém jazyce rekurentní neuronová síť zpracovává předpovědi sekvenčních dat nebo časových řad. Příkladem využití zmíněné sítě mohou být akciové a jiné finanční trhy, kdy na základě dat z minulosti dokáže predikovat budoucí chování. Tento typ sítě se taktéž využívá při překladech, kde dochází ke změně slovosledu v závislosti na přechodu z jednoho jazyka do druhého. Oproti dopředné neuronové síti mohou data v rámci skrytých vrstev putovat ve smyčkách a kruzích (vracet se z následujícího uzlu zpět do předchozího). V některých případech zahrnují tyto sítě i speciální skryté vrstvy tzv. kontextové vrstvy, které mají za úkol poskytovat zpětnou vazbu a zpřesňovat výstup. [17], [21]

3.1.5 Convolutional neural network

Convolutional neural networks (CNN) neboli konvoluční neuronová síť je inspirovaná na základě biologických procesů a používá se pro klasifikaci obrazu a detekci objektů. Konvoluční neuronové sítě jsou navrženy tak, aby pracovali s mřížkovými vstupy, které mají silné prostorové závislosti v určitých oblastech mřížky. Nejběžnějším příkladem dat strukturovaných do mřížky je 2D obrázek. Uvedená data vykazují prostorové závislosti, protože přilehlá prostorová místa v obrazech mají velmi často podobně zbarvené pixely. Oproti předchozím typům sítě se liší ve struktuře jednotlivých vrstev. V rámci této sítě je každá z vrstev trojrozměrná. Díky trojrozměrnosti vrstev je možné zachytit různé barvy, které vytvářejí obraz. Z toho důvodu u konvoluční neuronové sítě vzniká závislost založená na prostorových vzdálenostech. Hlavní výhodou při využívání těchto neuronových sítí je, že kromě využití pro rozpoznávání obrazových dat dokážou taktéž zpracovat všechna časová, prostorová a časoprostorová data. [17], [18]

3.2 Učení neuronových sítí

Učení neuronových sítí probíhá úpravou jednotlivých vah. V jednovrstvých neuronových sítích je učení relativně snadný proces, protože předpovídaná chyba (ztrátová funkce) může být vypočtena jako přímá funkční závislost jednotlivých vah, avšak u vícevrstvých sítí nastává problém v tom, že je ztrátová funkce počítána z vah v předchozích vrstvách. Učení tedy probíhá tak, že dochází k porovnání vypočtené (předpovídané) hodnoty se skutečnými vstupy a následně se neuronová síť snaží chybu minimalizovat. Jedním

z principů, jak toho co nejefektivněji dosáhnout, je backpropagation, jenž se hojně využívá u moderních neuronových sítí, které pracují se složitými vzorci, daty nebo vztahy.

3.2.1 Princip backpropagation

Některé neuronové sítě např. FFN – dopředné neuronové sítě zpracovávají data pouze jedním směrem (od vstupu po výstup). Během průchodu dat jsou jim v jednotlivých vrstvách za pomoci neuronů přidělovány různé váhy, které se předávají do další vrstvy a ta s nimi pracuje. Na výstupu jednotlivé váhy utvoří určitou funkci, a ta je následně porovnána se skutečnými hodnotami a dojde k vypočtení chyby, která může být využita pro další úpravy vah neuronové sítě.

Backpropagation funguje na zcela odlišném způsobu. Chybový signál, vznikající úpravou jednotlivých vah, může proudit oběma směry tedy od vstupu po výstup i opačně. Díky tomu je možné každý výpočet chyby, ke kterému dojde v následující vrstvě odeslat zpět do vrstvy předchozí, upravit váhu a znovu jej zpracovat. Neuronová síť následně porovnává, jak změna vah ovlivňuje celkovou chybu, kterou se snaží minimalizovat. Proces se opakuje několikrát během tréninkových cyklů a postupně zlepšuje schopnost sítě vytvářet přesně předpovědi. Tento proces však může být velmi citlivý na počáteční podmínky, rychlost učení i hloubku sítě. U rozsáhlejších sítí může docházet k tzv. explodujícímu nebo mizejícímu gradientu, kdy jsou vypočtené váhy příliš velké, nebo naopak malé, což může bránit efektivnímu učení. Problém může být vyřešen pomocí různých technik např. dávkovou normalizací, adaptivní rychlostí učení nebo strategiemi pro inicializaci vah. [17], [18], [22]

4 CHATGPT

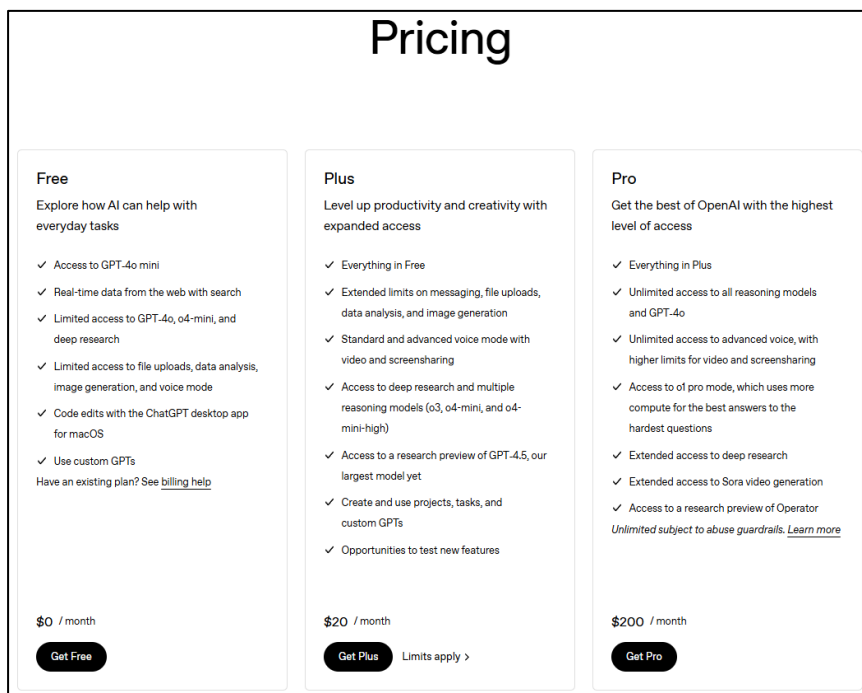
ChatGPT je jazykový model vytvořený společností OpenAI. Zkratka GPT znamená Generative Pre-trained Transformer neboli Generativní předtrénovaný transformátor. Jedná se o rozsáhlý jazykový model fungující na principu neuronové sítě. Zmíněný jazykový model rozumí lidským textům, dokáže je vyhodnotit a vést smysluplné konverzace. Díky těmto jedinečným dovednostem se ChatGPT stal nedílnou součástí firem i jednotlivců.

4.1 Společnost OpenAI

Společnost OpenAI byla založena koncem roku 2015 jako nezisková organizace fungující pod právní klasifikací 501(c)(3). Jejím cílem bylo naprogramovat bezpečnou a prospěšnou umělou inteligenci, která by byla lidem nápomocná. Vytvoření takto náročného technologického projektu vyžadovalo obrovské množství financí, které chtěla tato nezisková organizace získat v darech. Jejich prvotním cílem bylo získat 1 miliardu dolarů, která by zafinancovala provoz a počáteční výzkumné práce. V průběhu let se jim však podařilo získat pouze zlomek potřebné částky, konkrétně 130,5 milionu dolarů. Kvůli těmto skutečnostem bylo potřeba vymyslet jinou organizační strukturu. Nezisková organizace OpenAI zůstala zachována, ale v roce 2019 došlo k odštěpení dceřiné společnosti, která začala vydávat akcie za účelem zisku kapitálu. Nově vzniklá dceřiná společnost stále podléhala původnímu vedení neziskové organizace. Díky této organizační struktuře bylo možné zachovat důležité správní aspekty. Ve stejném roce došlo uzavření partnerství s Microsoftem, jenž do společnosti zainvestoval 1 miliardu dolarů. Partnerství spočívalo ve využívání cloudové infrastruktury Microsoftu společností OpenAI, což výrazně urychlilo a přispělo k trénování pokročilejších jazykových modelů. Dnes je společnost OpenAI technologickým lídrem v oblasti rozvoje jazykových modelů. [23]

4.2 Druhy předplatného

Pro běžné uživatele, kteří využívají jazykové modely přímo z prostřednictví webu společnosti OpenAI na stránkách <https://openai.com/chatgpt/pricing/>, je k dispozici pět různých cenových modelů, které obsahují odlišné verze chatbotů. Základní rozdíly v jednotlivých verzích jazykových modelů jsou popsány níže v kapitole „Verze jazykových modelů ChatGPT“.



Obrázek 3. Druhy předplatného [23]

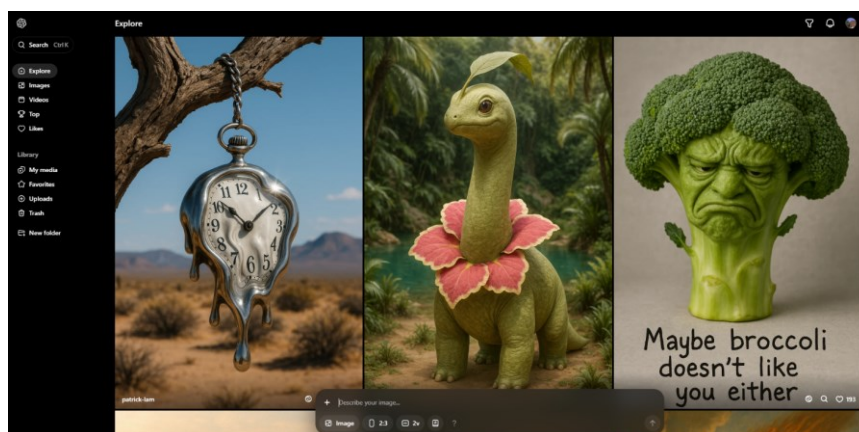
4.2.1 Free verze

První cenový model je zcela zdarma. Uživatel k němu získá přístup prostřednictvím registrace na webu společnosti. „Předplatné zdarma“ nabízí jazykový model GPT-4o mini, běžný hlasový režim, možnost nahrávat a analyzovat data ze souborů, obrázků a webových stránek. V omezené míře mají uživatelé v rámci pětihodinového okna přístup i k pokročilejšímu modelu GPT-4o, kam mohou odeslat omezený počet požadavků. Jakmile jsou tyto limity vyčerpány, je uživatel vyzván k upgradování svého uživatelského účtu na placenou verzi nebo k přechodu na zastaralejší verze jazykových modelů GPT-4o Mini nebo ChatGPT 3.5. Co se týče nástroje určeného ke generování obrázků DALL·E 3 mají uživatelé možnost generovat až tři obrázky denně.

4.2.2 Předplatné Plus

Předplatné „Plus“ je první z placených variant. V návaznosti na délku trvání předplatného a způsob jeho užívání se cenově mírně liší. Pokud je tento plán užíván jednotlivcem a placen měsíčně, stojí předplatné „Plus“ 20 dolarů. Tato cena může být snížena zakoupením ročního předplatného, které stojí 200 dolarů, čímž se cena za 1 měsíc sníží na 16,7 dolarů. Předplatné Plus obsahuje všechny funkce jako Free verze a zároveň umožňuje využívat GPT-4o bez omezení. Krom toho nabízí pokročilou analýzu dat, generování obrázků pomocí rozšiřujících nástrojů jako např. DALL·E 3 a omezené využití

nástroje Sora, který slouží ke generování videí pomocí promptů. Během měsíce je možné vytvořit až 50 videí v rozlišení 720p o délce 5 sekund. Uživatel si navíc může sám zvolit, jakou verzi jazykového modelu chce pro práci používat a v případě potřeby ji může kdykoliv změnit. Další velmi užitečnou funkcí je možnost hlasové konverzace s umělou inteligencí. Hlasový režim nabízí základní i pokročilou variantu. Základní hlasový režim dokáže vyslechnout uživatele a sdělit mu řešení jeho problému. Nevýhodou standartní varianty jsou občasné výpadky a horší přizpůsobení jazyku uživatele. Pokročilý hlasový režim dokáže plynule konverzovat a navázat na původní rozhovor i v případě, kdy uživatel chatbotovi skočí do řeči. V pokročilé variantě je možné libovolně přizpůsobit hlas a tonalitu umělé inteligence podle preferencí posluchače. Placená verze navíc umožňuje omezený přístup (80 zpráv každé 3 hodiny) k nejpokročilejším modelům o1 a o1-mini, kterým je věnována pozornost v následující kapitole této diplomové práce. Společnost OpenAI navíc své předplatitele využívá při testování nových jazykových modelů, díky čemuž se mohou seznámit s novinkami ještě dříve, než jsou volně k dispozici pro širší veřejnost. V poslední řadě je předplatitelům umožněno vytvořit a naučit na svých vlastních datech svého chatbota. Tato funkce je velmi užitečná pro firmy, e-shopy, vzdělávací či jiné platformy, protože takto natrénovaného chatbota lze implementovat na web pomocí API. [23], [24]



Obrázek 4. Sora – prostředí pro vytváření videí [23]

4.2.3 Předplatné Team

Předplatné Team je často využíváno menšími firmami, protože nabízí sdílené chaty v rámci oddělených přístupů. Stejně jako u předplatného Plus se měsíční cena liší v závislosti na délce předplatného. Cena měsíčního předplatného je nastavena na 30 dolarů za uživatele a v rámci ročního předplatného je cena nastavená na 300 dolarů, což je 25 dolarů měsíčně. Limit spolupracujících uživatelů není nijak omezen, nicméně společnost OpenAI na svém

fóru doporučuje, že při překročení 150 členů je vhodné kontaktovat jejich prodejní tým a ve spolupráci s tímto týmem zvážit přechod na nejvyšší stupeň Enterprise. Předplatné Team nabízí téměř totožné funkce jako Plus s rozdílem denních limitů, které jsou vyšší a možností spravovat firemní role a korigovat jednotlivé členy. Stejně jako u vyšší týmové verze Enterprise je firmám zaručeno, že jejich vstupní data nebudou využívána pro trénování nových jazykových modelů. [23], [24], [25]

4.2.4 Verze Pro

Varianta Pro je nejdražší a nepokročilejší uživatelskou variantou. Cena této varianty je 200 dolarů měsíčně a verze Pro se doporučuje převážně pokročilým uživatelům, kteří hledají specializované nástroje a potřebují co největší výpočetní výkon. Tato verze zahrnuje neomezený přístup ke všem dosavadním verzím jazykových modelů ChatGPT. Rovněž mohou uživatelé využívat model o1 v pro módu, který dokáže odpovídat na ještě složitější a komplikovanější dotazy než předchozí modely. Krom textových a obrázkových vstupů a výstupů je možné pracovat i s video formáty. Předplatitel získá přístup k rozšíření Sora, což je doprovodný plugin sloužící ke generování videí z textových promptů. Ve verzi Pro je uživatelům umožněno vytvořit až 500 prioritních videí měsíčně (jedná se o rychlé generování obsahu) a neomezený počet tzv. „relaxed videos“ – videa, jejichž tvorba zabere delší množství času, protože se výpočetní výkon rozkládá mezi všechny uživatele. Vytvořená videa mají rozlišení 1080p a délku 20 sekund. Navíc lze video stáhnout i bez vodoznaku společnosti nástroje Sora. [23], [24]

4.2.5 Předplatné Enterprise

Enterprise plní stejnou funkci jako předplatné Team. Slouží především firmám a podnikům, které využívají umělou inteligenci a potřebují mezi sebou sdílet jednotlivá pracovní prostředí. Enterprise nemá na webu uvedenou cenu, protože se liší v závislosti na velikosti a preferencích firmy, počtu přístupů a dalších faktorech. Pro zjištění cenové nabídky je nutné kontaktovat prodejní tým společnosti OpenAI. Ve srovnání s předplatným Team je vhodné pro více než 150 spolupracovníků a nabízí všechny nástroje bez jakéhokoli omezení podobně jako verze Pro. Stejně jako u verze Team je klientům zaručeno, že jejich vstupní data budou vyloučena z dalšího tréninku nových jazykových modelů. Dále lze v rámci této varianty ověřit doménu a pracovat s analytikami. [23]

4.3 Verze jazykových modelů ChatGPT

Společnost OpenAI v rámci svých balíčků a služeb nabízí několik verzí chatbota. Jednotlivé verze se od sebe liší ve schopnostech řešit zadané úlohy, výpočetním výkonu, které pro svou práci potřebují a stylem, jakým odpoví na zadané dotazy. Nejnovější verze dokonce před napsáním své odpovědi nad zadanou problematikou přemýšlí a vyhodnotí nejlepší možnou odpověď podobně jako lidé. [23], [26]

4.3.1 GPT 3.5

GPT 3.5 je označení pro starší jazykový model, který je stále k dispozici neplatícím uživatelům. V porovnání s dalšími verzemi je GPT 3.5 poměrně nepřesný a nespolehlivý v oblasti řešení složitých výpočetních úloh, logických úloh a řešení komplexních problémů. Vhodným využitím této verze je např. napojení pomocí API na webové rozhraní různých platforem a vytvoření chatbota, jenž zodpoví jednoduché dotazy týkající se produktů, či služeb. Dalším využitím může být práce s jednoduchým textem, kdy je GPT 3.5 použit ke shrnutí různých informací.

4.3.2 GPT-4o

Verze GPT-4o jazykového modelu společnosti OpenAI je všestranná, velmi inteligentní a vhodná ke každodennímu využití na běžných i složitějších úkolech. Přesto, že je jednou z nejběžněji užívaných verzí, samotné znalosti čerpá z dat, které mohla získat do října 2023. Po tomto datu byl ukončen výcvik a přijímání nových informací ukončen. Data je možné do uvedeného modelu vložit jak formou textu, tak formou obrázků. Stejně i výstup, který tento jazykový model vytvoří, může mít charakter textu i obrázků.

4.3.3 GPT-4o mini

Model GPT-4o mini se oproti verzi GPT-4o liší v rychlosti zpracování dat a cenové dostupnosti. Je vhodný spíše jako pomocník pro doladění a úpravu výstupů z větších modelů, protože zvládá především cílené úlohy. Stejně jako u komplexnějšího modelu GPT-4o byl výcvik a přijímání dat ukončen v říjnu roku 2023. [26]

4.3.4 Model o1 a o1-mini

Rozdíl modelů o1 a o1 – mini ve srovnání s předchozími modely spočívá v tom, že tyto modely mají vlastní paměť a předtím, než odpoví, „přemýšlejí“. Oba jazykové modely zvládají zpětnovazební učení, což jim umožňuje uvažovat o problémech mnohem

komplexněji a detailněji. Z dat uvedených na webu společnosti OpenAI plyne, že tyto modely vynikají ve vědeckém uvažování a předmětech jako matematika, fyzika, biologie, chemie a programování. V rámci zmíněných předmětů dokonce v některých oblastech překonaly znalosti doktorandů.

Model o1 je navržen tak, aby dokázal přemýšlet nad velmi složitými otázkami, přičemž využívá dosavadní znalosti o světě a jeho fungování. V porovnání s modelem o1 je model o1-mini cenově dostupnější a vhodnější pro řešení konkrétních úloh, ve kterých není potřeba vycházet z rozsáhlých obecných znalostí. Uživatelům může být nápomocný při programování, matematických výpočtech, řešení otázek spojených s fyzikou a přírodními vědami. [24], [26]

5 HODNOCENÍ DAT NA STATISTICKÉ BÁZI

5.1 Metodika EDA

Explorační datová analýza neboli EDA, je statistická metoda založena na vizualizaci a grafickém posouzení naměřených dat. V praxi je totiž velmi časté, že naměřená data nepodléhají normálnímu rozdělení, nemají konstantní rozptyl a porušují obecné předpoklady, které jsou pro další zpracování nezbytné. Z tohoto důvodu se metodika EDA využívá jako první krok při hlubším zkoumání vlastností výběrových souborů. Prvním krokem při provádění této datové analýzy je využití interaktivní statistické analýzy. Většina statistických softwarů umožňuje vytvářet různé grafy a diagramy, kterými lze odhalit např. přítomnost podezřelých hodnot, tvarové zvláštnosti, celkové rozdělení a rozprostření dat, normalitu (pomocí Gaussova normálního rozdělení) a jiné zvláštnosti. [27]

5.1.1 Určení typu dat

Ve statistice je možné data rozdělit podle jejich typu a toho, zda představují numerické hodnoty, nebo naopak slouží k vyjádření nějakých popisů. Podle těchto kritérií lze dělit data na kvantitativní a kvalitativní.

- Kvantitativní data představují číselné (numerické) hodnoty, které je možné dále zpracovávat a jsou vyjádřena číselně. Kvantitativní data se nadále dělí na diskrétní a spojitá. Diskrétní data jsou vyjádřena celými spočtenými hodnotami např. počty studentů ve škole, počty prodaného zboží v obchodě atd. Spojitá data mohou nabývat libovolných hodnot a často se měří s určitou přesností. Může se jednat například o výšku člověka (185,7 cm), hmotnost (96,2 kg) atd.
- Kvalitativní data vyjadřují spíše vlastnosti či kategorie, podle kterých lze jednotlivé hodnoty seřadit či porovnat. Tento druh dat je reprezentován textem a může být rozdělen do následujících kategorií – binární data, nominální data a ordinální data. Binární data jsou zastoupena pouze dvěma možnostmi 1/0, pravda / nepravda, ano / ne. Nominální data jsou kategorie bez přirozeného pořadí. Není možné je žádným způsobem uspořádat a jedná se například o barvu očí, národnost, pohlaví atd. Ordinální data mohou mít své přirozené pořadí a rozdíly mezi nimi nemusí být stejné. Jedná se o pořadí (1., 2., 3. místo), dosažené vzdělání (základní, střední, vysokoškolské ...) [27]

5.1.2 Grafické zobrazení hodnot

V závislosti na softwaru se data zobrazují pomocí nejrůznějších typů grafů. Jednotlivá zobrazení mohou pomoci při odhalení různých vlastností výběrových souborů a analýze dat. Níže zvolené grafy jsou pouze výčtem z mnoha dalších.

- Časová řada – využívá se pro zobrazení naměřených hodnot v čase. Díky tomu, že jsou hodnoty zobrazeny tak, jak došlo k jejich naměření umožňuje identifikovat trendy výkyvy v datech.
- Diagram rozptýlení – tento diagram slouží především pro kontrolu rozptýlení naměřených hodnot a dokáže identifikovat shluky a podezřelé či odlehlé hodnoty.
- Krabicový graf – slouží stejně jako předchozí grafy k vizualizaci naměřených hodnot a lze z něj vyčíst rozdělení dat a podezřelé či odlehlé hodnoty. Krabicový graf navíc zobrazuje minimální a maximální hodnotu, medián a jednotlivé kvartily výběrového souboru.
- Histogram – jedná se o typ sloupcového grafu, kde je na ose x zobrazena šířka jednotlivých sloupců a jejich výška následně odpovídá empirickým hustotám pravděpodobností. Histogram umožňuje analyzovat četnost výskytu hodnot ve výběrovém souboru dat a pomáhá odhalit tvar rozložení.
- Kvantilový graf – díky své přehlednosti umožňuje rozlišit tvar rozdělení hodnot a dokáže odhalit, zda je jejich uspořádání symetrické či případně dochází k zešikmení. Tento graf tedy pomáhá identifikovat odchylky od teoretického modelu normálního rozdělení a umožňuje odhalit asymetrii a přítomnost extrémních hodnot.

5.1.3 Metody vyjádření a interpretace dat

Po provedení metody EDA je dalším nezbytným krokem správná interpretace dat, protože může poskytnout hlubší porozumění jejich jednotlivým charakteristikám a vlastnostem. U kvantitativních dat lze vyjádřit aritmetický průměr, rozptyl, směrodatnou odchylku a variační koeficient. Dále je možné určit medián, horní a dolní kvartil či maximální a minimální hodnotu. Kromě těchto základních datových charakteristik je dobré prošetřit, zda mají data normální rozdělení, s čímž může pomoci Q-Q diagram, který byl zmíněn výše, Shapirův-Wikův nebo Kolmogorovův-Smirnovův test. Velmi důležitým

aspektem je taktéž maximální a minimální hodnota, jež je dobré prověřit, zda se náhodou nejedná o odlehlou hodnotu, kterou může odhalit např. Grubbsův test odlehlých hodnot. Statistika nabízí spoustu dalších velmi užitečných nástrojů určených ke zkoumání jednotlivých charakteristik výběrových souborů a jejich využití se odvíjí od toho, co je potřeba pro různé účely analyzovat. [27], [28]

5.2 Regresní analýza

Regresní analýzu lze zařadit do statistických metod, které se zabývají zkoumáním vztahů mezi dvěma a více proměnnými. Hlavním cílem této analýzy je identifikace vztahů mezi závislou a nezávislou proměnnou, na základě nichž je možné sestavit matematický model, předpovídající hodnoty závislé proměnné z hodnot nezávislé proměnné. Vytváření takového modelu se skládá z několika hlavních kroků, kterými je určení lineární rovnice a jejího polynomu, odhad jednotlivých parametrů, hodnocení modelu, testování normality reziduí a vyhodnocení výsledků. V praxi nachází regresní analýza široké využití zejména v oblastech ekonomie, inženýrství nebo umělé inteligence. [28]

5.2.1 Pravidla pro sestavení lineární regrese

Lineární regrese patří k jedné z nejpoužívanějších metod sloužících k prediktivní analýze. Správné sestavení je klíčovým aspektem tohoto nástroje, protože nesprávně určená regrese může vést k chybným výsledkům. V této kapitole diplomové bude nahlédnuto na proces správného sestavení lineární regrese.

- Určení závislé a nezávislé proměnné je prvním a zcela zásadním krokem.
- Následný krok zahrnuje stanovení správného pracovního rozsahu – nezávislá proměnná a rozsahu výsledků – závislá proměnná.
- Nalezení vhodných regresních parametrů podle kritéria optimality. V praxi je nejčastěji využívána metoda nejmenších čtverců.
- Statistická analýza reziduí – je zcela nezbytné, aby byla rezidua zatížena pouze náhodnou chybou (nesmí dojít k porušení centrálního limitního teorému). K ověření normality dat lze využít test normality a k analýze odlehlých či podezřelých hodnot W – graf.

- Prověření jednotlivých regresních koeficientů. Každý regresní koeficient musí být statisticky významný tzn. p – hodnota je menší než hladina významnosti α . V opačném případě je model nevýznamný a může být zbytečně overfitted. [27], [28]

6 HODNOCENÍ DAT NA STATISTICKÉ BÁZI

6.1 Zdroje vzniku a typy chyb při měření

Při každém měření mohou vznikat určité chyby, za nimiž může stát několik faktorů. Tato kapitola diplomové práce objasňuje základní faktory, kterými mohou být chyby jako takové způsobeny. Obecně lze chyby rozdělit podle následujících kategorií.

6.1.1 Chyby při měření

Již při samotném měření může dojít ke vzniku několika druhů chyb, které jsou způsobeny konstrukcí měřidla, použitou metodou měření, zpracováním naměřených dat nebo užitím špatného postupu měření. Chyby spojené s konstrukcí měřidla jsou ve většině případů uváděny výrobcem a souvisejí se samotnou konstrukcí měřidla nebo měřícího přístroje. Metodické chyby vznikají během měření a mohou nastat špatným odečtem rozměrů z měřícího přístroje, vnějšími vlivy, jako jsou teplotní výkyvy, vibrace anebo nevhodnými podmínkami při měření. Dalším druhem chyb, který může nastat, jsou chyby teoretické, spojené s určitým postupem měření. Mohou vznikat v závislosti na fyzikálních modelech, použitých parametrech nebo kvůli využití fyzikálních konstant. Ve shrnutí se dá říci, že tyto chyby nevznikají náhodným vlivem, ani nejsou způsobeny přístroji, protože je způsobuje využití různých teoretických předpokladů, které zcela neodpovídají realitě a v důsledku toho mohou být ve výsledku nepřesné. [27]

6.1.2 Chyba náhodná, systematická a hrubá

Chyby lze dále klasifikovat do několika kategorií podle příčin jejich vzniku.

Náhodná chyba vzniká zcela náhodně během opakování při měření. Tato chyba je nepředvídatelná a za jejím vznikem může být mnoho faktorů, které nelze žádným způsobem odstranit či omezit.

Systematická chyba se ve srovnání s náhodnou chybou liší ve způsobu jejího vzniku, protože vzniká v závislosti na určitých faktorech. Jedním z faktorů může být např. špatné nastavení nulové hodnoty měřícího přístroje. Naměřená veličina se oproti skutečné veličině liší pokaždé o stejnou velikost. Tuto chybu je možné odstranit korekcí – odečtením velikosti systematické chyby od naměřené hodnoty.

Za hrubou chybu lze považovat takovou hodnotu, která se od ostatních naměřených hodnot liší výrazným způsobem a do datové řady zcela nezapadá. Hrubá chyba bývá

způsobena především náhlým selháním měřicí aparatury, špatně zaznamenaným výsledkem či jinými vlivy. [27]

6.2 Vliv chyb na měření pomocí laserových interferometrů

V průmyslovém odvětví se dá laserová interferometrie považovat za jednu z nejpresnějších metod měření polohy. Stejně jako u každé metody může být i měření pomocí laserové interferometrie ovlivněno řadou negativních vlivů a chyb, majících zásadní dopad na výsledek měření. V této kapitole je poukázáno na nejčastější chyby, které se během tohoto typu měření mohou vyskytnout. [9]

6.2.1 Změna teploty

Změnou teploty dochází u různých materiálů k různé tepelné roztažnosti, která dokáže ovlivnit přesnost měření. Nejčastěji dochází ke změně délky optických vláken, což ve výsledku může způsobit nepřesnosti. [29]

6.2.2 Vibrace

Dalším velmi zásadním faktorem ovlivňujícím přesnost měření pomocí interferometru jsou vibrace. Interferometry jsou velmi citlivé a jakékoliv vibrace mohou vést k nežádoucím interferenčním jevům. [13]

6.2.3 Optické chyby a aberace

Optické vady jako nesférické čočky nebo odrazové artefakty mohou taktéž ovlivnit měření negativním způsobem. [10]

6.2.4 Interferenční efekty a chybové signály

Posledním z vlivů, kterému je v této práci věnována pozornost, jsou parazitické rušivé vlivy. Zdrojem těchto chyb může být různá vlnová délka světla použitého u měření. [29]

6.3 Kalibrace robotů

Kalibrace výrobních robotů je proces sloužící ke zlepšení absolutní přesnosti polohování průmyslových manipulátorů. Díky tomuto procesu je možné minimalizovat systematické chyby vznikající při výrobě. Stejně jako jiná mechanická zařízení pracují výrobní roboti s určitou přesností, která je dána nejen výrobní tolerancí a použitými součástkami, ale i opotřebením a dalšími vlivy. Z uvedených důvodů je nezbytné u

výrobních robotů provádět kalibrace, pomáhající při eliminaci systematických chyb. Kalibrací dochází ke včasnému detekování nástupu systematických vlivů a jeho následnému odstranění pomocí úprav řídicího softwaru robota. [4]

6.3.1 Proces provádění kalibrace

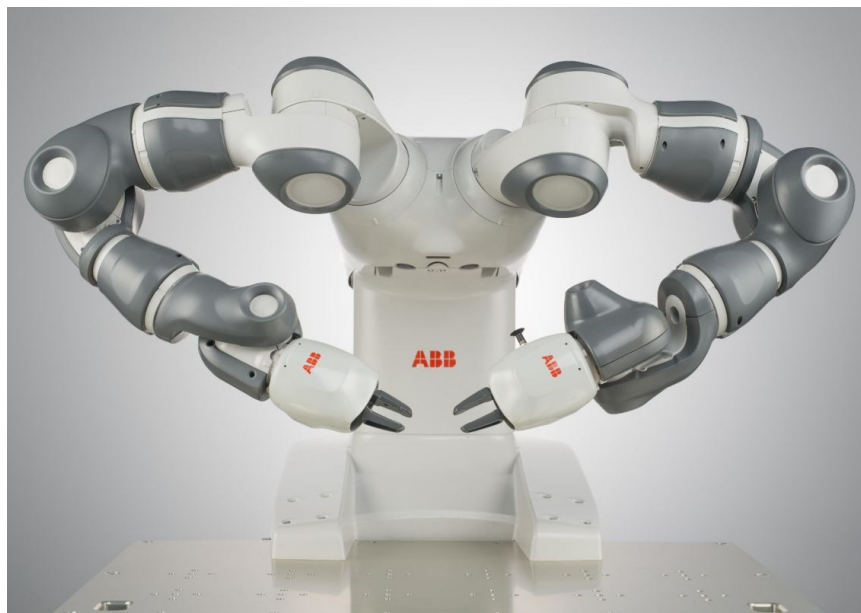
Ve většině případů dochází během kalibrace k porovnání mezi skutečnou polohou robota, kterou je možné určit např. pomocí laserového interferometru a modelovými hodnotami. Proces kalibrace spočívá v určení odchylek, které se vypočítají jako rozdíl mezi skutečnou a modelovou polohou. Následně dochází k upravení kinematických parametrů podle zjištěných rozdílů. Tímto způsobem je možné snížit absolutní chybu a zvýšit přesnost. [4]

7 ROBOTIKA

Robotika jako taková je vědní disciplína zabývající se designem, výrobou a aplikací robotů např. v průmyslu. Nasazení robotů do průmyslu umožňuje zvýšení efektivity a přesnosti výrobních procesů. V současnosti jsou průmysloví roboti běžně využíváni pro řadu technologických operací, jako je manipulace, montáž, svařování a mnoho dalších. Automatizací výroby pomocí robotů dochází ke snížení výrobních časů a chybovosti. Významnou roli v této oblasti hrají číslíkově řízené stroje (CNC), které umožňují velmi přesnou manipulaci s výrobky a efektivní řízení procesu. [30], [31]

7.1 Průmyslový robot IRB 14000 - YuMi

V roce 2015 společnost ABB představila kolaborativního robota IRB 14000 pod přezdívkou „YuMi“. Zkratka YuMi vznikla ze slov „You“ a „Me“ v překladu „ty“ a „já“, protože cílem společnosti ABB je, aby lidé a roboti dokázali spolupracovat na stejných úkolech při zajištění dostatečné bezpečnosti. IRB 14000 - YuMi je průmyslový robot se dvěma pažemi, který se používá pro montáž drobných součástek. Výhodou tohoto průmyslového robota je, že dokáže spolupracovat s lidmi v běžném výrobním prostředí bez toho, aniž by byla ohrožena jejich bezpečnost.



Obrázek 5. Průmyslový robot IRB 14000 – YuMi [32]

7.2 Popis robota Yumi z hlediska jeho kalibrace

Dle manuálu společnosti ABB by měl být robot YuMi kalibrován v následujících případech:

- pokud dojde ke změně hodnot resolveru;
- při ztrátě paměti otáčkoměru;
- při přestavbě robota.

7.2.1 Kalibrace při změně hodnot resolveru

Resolver je rotační typ elektrického transformátoru, který se používá pro měření stupňů rotace. Pokud dojde ke změně hodnot tohoto zařízení, je nutné provést opětovnou kalibraci robota YuMi. Tuto kalibraci je potřeba provádět bez namontovaného nástroje a spočívá v uvolnění přídržných brzd jednotlivých os. Následně se ručně pohne ramenem tak, aby došlo zarovnání jednotlivých synchronizačních značek. Po zarovnání kloubu je nutné opětovně zapnout rychlosti jednotlivých motorů na 100 % a v programu FlexPendantu vybrat možnost jemné kalibrace s hallovými senzory CalHall. Potvrzením bude uživatel vyzván k tomu, aby vybral klouby, které chce jemně kalibrovat. Kliknutím na tlačítko „OK“ dojde k zahájení jemné kalibrace, během níž se bude vybraný kloub několikrát otáčet tam a zpět, aby došlo k nalezení přesné polohy, ve které je hallův senzor zarovnán. Po provedení kalibrace je možné znovu přimontovat jednotlivé nástroje, které musely být pro kalibraci odstraněny.

7.2.2 Kalibrace při ztrátě paměti otáčkoměru

Ke ztrátě paměti otáčkoměru může dojít vybitím baterie, chybě resolveru, přerušením signálu mezi resolverem a měřicí deskou, nebo jakýmkoliv pohybem robota, který má odpojený řídicí systém. Pokud dojde k některé z těchto situací, je nutné aktualizovat čítač otáček. Čítač otáček je potřeba aktualizovat i v případě prvního připojení robota a řídicí desky při první instalaci. Hrubou kalibraci každé osy robota je možné provést následovně. Stejně jako u předchozí kalibrace je nutné ručně přesunout manipulátor do kalibrační polohy, v programu vybrat kalibraci pomocí Hallových senzorů CalHall a vybrat funkci „Aktualizovat čítače otáček“.

7.2.3 Při přestavbě robota

Pokud z nějakého důvodu došlo k přestavění robota např. vlivem havárie, výměnou senzorů nebo jiným způsobem, je důležité ho znovu zkalibrovat. Kalibrace se v tomto případě řídí pravidly popsány v následující kapitole. [32]

7.3 Kalibrace robotických paží

Kalibrace robotických paží je nezbytná pro zachování přesnosti robotů, protože výrobní roboti velmi často vykonávají stejný pohyb a následně se vrací zpět do výchozí polohy. Při kalibraci dochází k porovnání mezi požadovanou a naměřenou hodnotou. Rozdíl mezi těmito hodnotami může být způsoben výrobními tolerancemi robota, přesností montáže, opotřebením a dalšími vlivy. Kalibrací robotických paží lze zvýšit přesnost a minimalizovat systematické chyby, vznikající jako následek opakovatelnosti.

7.3.1 CalibWare

Jednou z možných metod kalibrace robotických ramen je metoda CalibWare. Jedná se o standardní metodu, na kterou je potřeba externí zařízení jako laserový interferometr. Během měření dochází ke zjištění skutečné polohy robotické paže a porovnání s teoretickou hodnotou. Rozdíl mezi těmito hodnotami se zadá do softwaru a tím dojde ke kompenzaci odchylek. K naměření skutečných hodnot robotických paží lze využít laserové interferometrie, která je díky své přesnosti jednou z nejvhodnějších metod. Při měření odchylek robotických paží může docházet vlivem pohybů robota ke vzniku vibrací, jenž mohou negativně ovlivnit celé měření. Z tohoto důvodu je nezbytné využít speciální přípravky či jiné prvky, zajišťující dostatečnou stabilitu a snižující negativní vlivy. [31], [32], [33]

7.3.2 YuMi AbsAcc Recovery

Druhý způsob kalibrace robotických paží je metoda YuMi AbsAcc Recovery, kterou lze využít pouze u robotů s verzí RobotWare 6.05.01 či novější. Ve srovnání s předchozí metodou není potřeba žádné další externí zařízení ani výpočet. YuMi AbsAcc Recovery je možné využít pouze tehdy, pokud došlo k výměně jednoho ramene. K určení kalibračních hodnot dochází za pomoci dat z druhého již seřazeného ramene pomocí přípravků, který společnost ABB nabízí na svých webových stránkách. Ještě před spojením ramen pomocí přípravku je potřeba provést jednou kalibraci neseřazeného ramene. Postup provádění jemné kalibrace je zmíněn na začátku této kapitoly. Po provedení jemné kalibrace je možné spojit

ramena a zahájit kalibraci. Nevýhodou tohoto způsobu je, že přesnost kalibrace závisí na seřazeném rameni. Pokud by došlo ve větším časovém horizontu k několika výměnám obou ramen, může dojít ke ztrátě přesnosti, protože tato metoda nedokáže plně nahradit přesnost CalibWare. [32]



Obrázek 6. Kalibrace robota pomocí spojených paží [32]

II PRAKTICKÁ ČÁST

8 CÍLE PRÁCE

Cílem diplomové práce je zpracovat naměřená data z hlediska regresní analýzy a naléznout nejvhodnější regresní rovnici. Na základě těchto poznatků vycvičit jazykový model společnosti OpenAI – ChatGPT 4o tak, aby data zpracoval a vyhodnotil stejným způsobem. Výsledkem celé práce bude optimalizovaný prompt – program, díky němuž bude možné provést regresní analýzu bez hlubších znalostí statistiky. Hlavní výhodou této metody je rychlost, efektivita a úspora peněz, protože uživatel k vyhodnocení potřebuje pouze internet a předplacenou verzi ChatGPT 4o. Díky tomu není potřeba žádný dodatečný software, ani licence, které se dnes pohybují v řádech desetitisíců korun.

9 EXPERIMENT

9.1 Měření dat

Měření a získávání dat probíhalo v laboratořích Univerzity Tomáše Bati ve Zlíně v budově Fakulty technologické. V rámci experimentu byl využit průmyslový robot YuMi a laser interferometr Renishaw XL – 80. Interferometrické odražeče byly pomocí speciálních gripů umístěny do robotických paží průmyslového robota a měření spočívalo v polohování paže do předem definovaných pozic. Celá měřicí sestava včetně uložení odražečů je zachycena na obrázku níže.



Obrázek 7. Průběh měření a získávání dat

(zdroj: vlastní)

9.2 Zpracování naměřených dat

Měření probíhalo v šesti cyklech, přičemž ve startovací pozici byla robotická paže ve vzdálenosti 60 mm od výchozího bodu a následně se vzdalovala vždy o 60 mm až do vzdálenosti 480 mm. Následně se paže vracela po stejných vzdálenostech zpět do výchozího bodu. Interferometr zaznamenával odchylku skutečné pozice od nominální hodnoty.

9.2.1 Explorační analýza dat – EDA

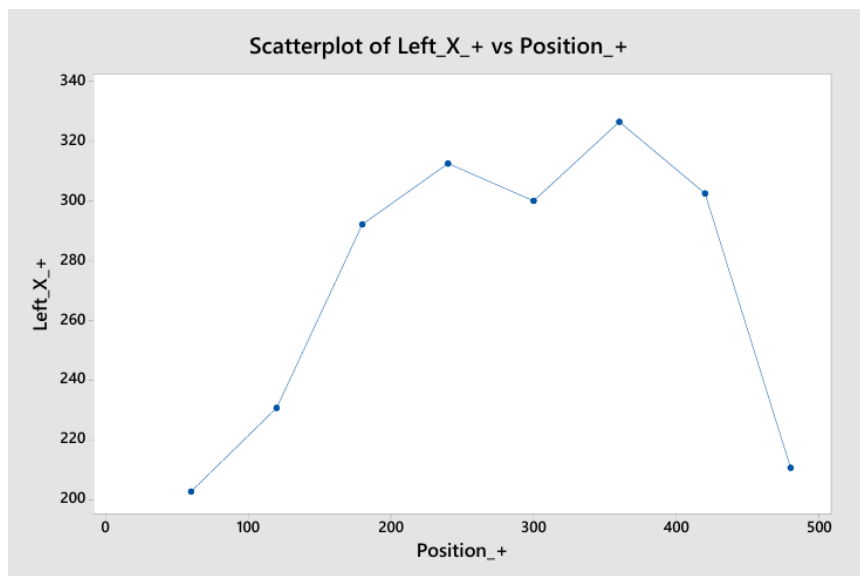
Jako první byla naměřená data podrobena metodice EDA – Explorační datové analýze, přičemž bylo hlavním cílem zjistit, zda data podléhají normálnímu rozdělení a nenachází se v nich odlehlé hodnoty či hrubé chyby. Celá analýza byla provedena v softwaru Minitab 17. Naměřená data podléhala pouze náhodnému rozdělení a žádné odlehlé hodnoty ani hrubé chyby neobsahovala. Následně došlo k vypočítání jednotlivých středních hodnot v rámci všech pozic a směrodatných odchylek.

Tabulka 1. Střední hodnoty odchylek v jednotlivých pozicích

Pozice	Střední hodnoty odchylek [μm]	Směrodatná odchylka odchylek [μm]
60	202,667	12,675
120	230,667	17,478
180	292,167	43,032
240	312,500	34,069
300	300,000	54,391
360	326,500	42,578
420	302,500	40,894
480	210,667	37,814

(zdroj: vlastní)

Zajímavostí je, že na níže uvedeném grafu rozptylů lze pozorovat výrazný nárůst směrodatných odchylek v oblasti pozic 180 mm a 420 mm. Jedná se o rozsah pohybů, při kterých bylo robotické rameno s největší pravděpodobností využíváno nejvíce. Z toho důvodu došlo k výraznějšímu opotřebení kloubů, a to zapříčinilo vyšší směrodatné odchylky.



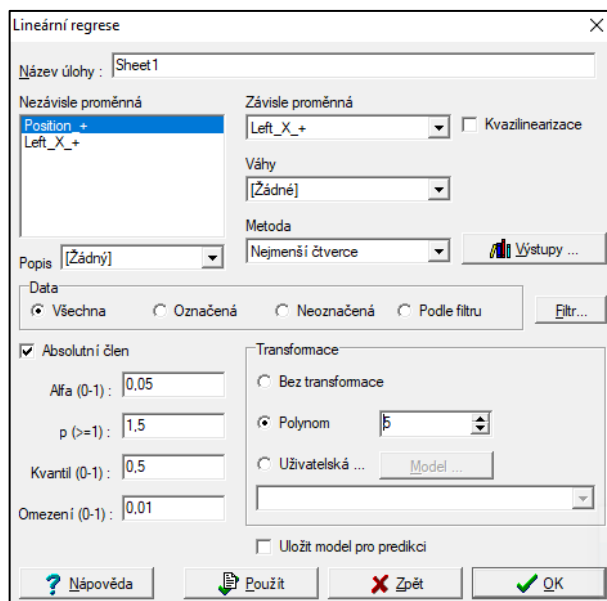
Obrázek 8. Graf rozptylů středních hodnot v jednotlivých pozicích

(zdroj: vlastní)

Takto připravená data je možné použít pro provedení regresní analýzy.

9.2.2 Regresní analýza

Regresní analýza byla provedena ve statistickém softwaru QC.Expert. Cílem regresní analýzy je nalézt funkční vztah mezi nastavenou pozicí a zjištěnou odchylkou. Testovalo se několik regresních modelů – lineární regrese 1. – 5. stupně. Jako první byla vypočtena hodnota Pearsonova korelačního koeficientu, která indikovala silnou korelaci. Dále došlo k určení závislé a nezávislé proměnné, přičemž za závislou proměnou byla zvolena střední hodnota odchylek a jako nezávislá proměnná nominální poloha robota. Na obrázku níže je zaznamenáno nastavení softwaru QC.Expert před spuštěním regresní analýzy pro polynom 5. stupně.



Obrázek 9. Nastavení Lineární regrese v programu QC.Expert

(zdroj: vlastní)

Dle výše nastavených parametrů byla provedena regresní analýza pro polynom 5. stupně. Regresní modely jsou hodnoceny podle následujících statistických parametrů:

- P – hodnoty odhadů regresních parametrů představují nejdůležitější hodnoty, které určují, zda jsou jednotlivé parametry statisticky významné, či nevýznamné. Přítomnost nevýznamných parametrů může vést k přeučení modelu tzv. overfitted, kdy je model místo skutečných hodnot přizpůsoben náhodnému šumu.
- Střední kvadratická chyba predikce – MEP ukazuje přesnost predikce modelu.
- Koeficient determinace R^2 nabývá maximální hodnoty 1 a udává, jaký podíl variability závisle proměnné model vysvětluje. V aktuálním případě je koeficient determinace 0,978, což znamená, že má regresní model vynikající shodu s reálnými daty.
- Akaikeho informační kritérium AIC – slouží k porovnání regresních modelů s ohledem na jejich složitost.

Analýza rozptylu						
Průměr Y :	272,2085					
Zdroj	Součet čtverců	Průměrný čtverec	Rozptyl			
Celková variabilita	17008,32878	2126,041097	2429,761254			
Variabilita vysvětlená modelem	16637,53117	2079,691397	2376,790168			
Reziduální variabilita	370,7976126	46,34970157	52,97108751			
Hodnota kritéria F :	17,94782987					
Kvantil F (1-alfa, m-1, n-m) :	19,29640965					
Pravděpodobnost :	0,05361445318					
Závěr :	Model je nevýznamný					
Odhady parametrů						
Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	349,583	140,3517836	Nevýznamný	0,1303946375	-254,3019849	953,4679849
Position_+	-5,439937942	4,112856621	Nevýznamný	0,3169275412	-23,13613171	12,25625583
Position_+2	0,06593880932	0,04073500758	Nevýznamný	0,2469228604	-0,1093297822	0,2412074009
Position_+3	-0,0003082060576	0,0001782786054	Nevýznamný	0,2259878454	-0,001075276986	0,0004588648704
Position_+4	6,398703694E-007	3,549410511E-007	Nevýznamný	0,2132082877	-8,873177132E-007	2,167058452E-006
Position_+5	-4,946177258E-010	2,622822725E-010	Nevýznamný	0,1999688311	-1,623127261E-009	6,338918098E-010
Statistické charakteristiky regrese						
Vícenásobný korelační koeficient R :	0,9890394616					
Koeficient determinace R ² :	0,9781990566					
Predikovaný korelační koeficient Rp :	37,15894098					
Střední kvadratická chyba predikce MEP :	15085,9909					
Akaikeho informační kritérium :	42,68971883					

Obrázek 10. Výsledek regresní analýzy pro polynom 5. stupně

(zdroj: vlastní)

Z výsledků této analýzy vyplývá, že je statisticky nevýznamný. Jednotlivé odhady regresních parametrů jsou taktéž nevýznamné a z uvedeného důvodu bude potřeba spustit analýzu znovu s jiným polynomickým stupněm.

V rámci testování došlo k provedení jednotlivých regresních analýz od 5. do 1. stupně a jako jediný model, jenž měl všechny parametry statisticky významné, se ukázal 2. stupeň.

Analýza rozptylu						
Průměr Y :	272,2085					
Zdroj	Součet čtverců	Průměrný čtverec	Rozptyl			
Celková variabilita	17008,32878	2126,041097	2429,761254			
Variabilita vysvětlená modelem	14772,31039	1846,538798	2110,330055			
Reziduální variabilita	2236,018392	279,502299	319,4311888			
Hodnota kritéria F :	16,51631136					
Kvantil F (1-alfa, m-1, n-m) :	5,786135043					
Pravděpodobnost :	0,006266633883					
Závěr :	Model je významný					
Odhady parametrů						
Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	111,9054286	29,50338992	Významný	0,0127196892	36,06455036	187,7463068
Position_+	1,432469444	0,250701109	Významný	0,002294799331	0,7880217274	2,076917161
Position_+2	-0,002466924603	0,0004532058863	Významný	0,002841318953	-0,003631927422	-0,001301921784
Statistické charakteristiky regrese						
Vícenásobný korelační koeficient R :	0,93195167					
Koeficient determinace R ² :	0,8885339153					
Predikovaný korelační koeficient Rp :	0,1612068481					
Střední kvadratická chyba predikce MEP :	1272,423423					
Akaikeho informační kritérium :	51,06408414					

Obrázek 11. Výsledek regresní analýzy pro polynom 5. stupně

(zdroj: vlastní)

Dalším cílem praktické části této diplomové práce bude naprogramovat umělou inteligenci tak, aby dokázala na základě vložených dat závislé a nezávislé proměnné provést regresní analýzu a dospět ke stejným výsledkům jako při práci se statistickým softwarem QC.Expert.

9.3 Vytvoření promptu pomocí ChatGPT

Experiment se odehrával v období od 1.1.2025 do 9.3.2025 a využíval verzi ChatGPT 4o jazykového modelu společnosti OpenAI. Jako první bylo potřeba definovat kostru samotného promptu (programu). Jedná se o přesně definovaný seznam úkolů, které musí ChatGPT vykonat, aby se dopracoval k požadovanému výsledku. Jakmile nějaký z kroků chybí, nebo je zadán špatně, dochází ke špatnému zpracování a vyhodnocení dat. Na základě těchto výstupů je prompt průběžně upravován do doby, než ChatGPT vyhodnotí výsledky požadovaným způsobem. Optimalizovaný prompt následně poslouží uživatelům, kteří nemají hluboké znalosti statistiky a pomůže jim vyhodnotit naměřená data z hlediska vytváření regresní analýzy.

9.3.1 První experiment

Základ skutečného promptu vytvořil jednoduchý experiment, který zkoumal, jakým způsobem ChatGPT reaguje na zcela otevřené zadání. Cílem bylo zjistit, jakým způsobem dochází ke zpracování dat a jak jazykový model uvažuje při jejich zpracování. Pro lepší orientaci v celé konverzaci budou vstupy označeny nadpisem „Vstup“ a výstupy nadpisem „Výstup“. Výstupy budou zároveň zvýrazněny kurzívou.

Vstup

Nalezni regresní rovnici maximálně 3. stupně, kde nezávisle proměnná bude position_+ a závisle proměnná je left_x_+ Position_+ left_x_+ 60 202,667 120 230,667 180 292,167 240 312,5 300 300 360 326,5 420 302,5 480 210,667

Výstup

Regresní rovnice maximálně 3. stupně je:

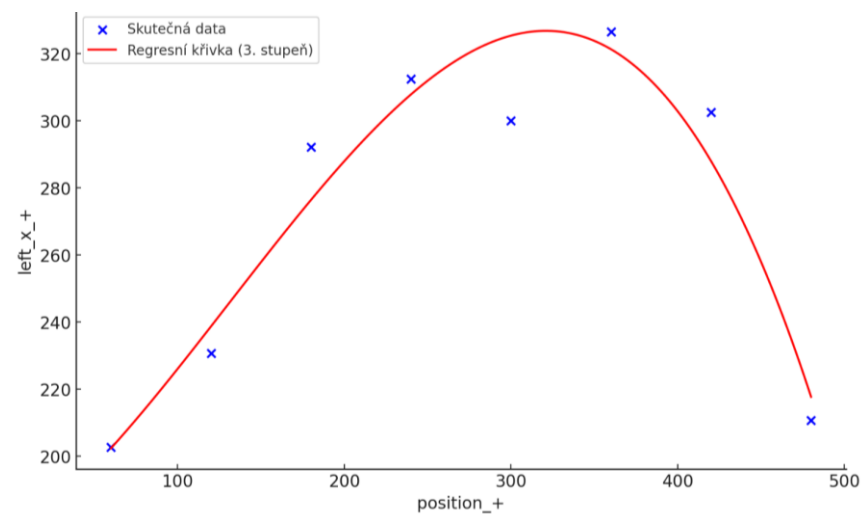
$$\text{left_x_+} = 175.155 + (0.336 * \text{position_+}^1) + (0.002 * \text{position_+}^2) - (0.000 * \text{position_+}^3)$$

Tento model popisuje závislost mezi proměnnou left_x_+ a position_+. Pokud budeš chtít vizualizaci nebo ověření kvality modelu, dej vědět.

Vstup

Ano, přidej i vizualizaci.

Výstup



Obrázek 12. Vizualizace první regresní analýzy pomocí ChatGPT [34]

Tady je vizualizace regresní analýzy. Modré body reprezentují skutečná data, zatímco červená křivka je modelovaná regresní rovnice třetího stupně. Pokud bys potřeboval další analýzu nebo úpravy, dej vědět!

Vstup

Proveď výpočet statistické významnosti parametrů regresní rovnice position_+ a dle nich uprav regresní rovnici.

Výstup

Statistická analýza ukazuje, že pouze konstantní člen (intercept) je statisticky významný ($p < 0.05$). Parametry pro vyšší stupně proměnné position_+ nejsou statisticky významné. Nová regresní rovnice tedy bude zjednodušená pouze na konstantní člen: $\text{left_x_+} = 272.208$

Pokud bys chtěl další podrobnosti nebo jiný přístup, dej vědět!

Vstup

Uprav regresní rovnici na polynom 2. stupně a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05

Výstup

Regresní analýza s polynomem 2. stupně ukazuje následující výsledky:

p-hodnoty pro jednotlivé koeficienty jsou všechny menší než 0,05, což znamená, že všechny parametry jsou statisticky významné.

Regresní rovnice 2. stupně je tedy:

$$\text{left_x_+} = 111.905 + (1.432 * \text{position_+}) + (-0.002 * \text{position_+}^2)$$

Pokud bys chtěl další analýzy nebo vizualizaci této rovnice, dej vědět!

Závěr prvního experimentu

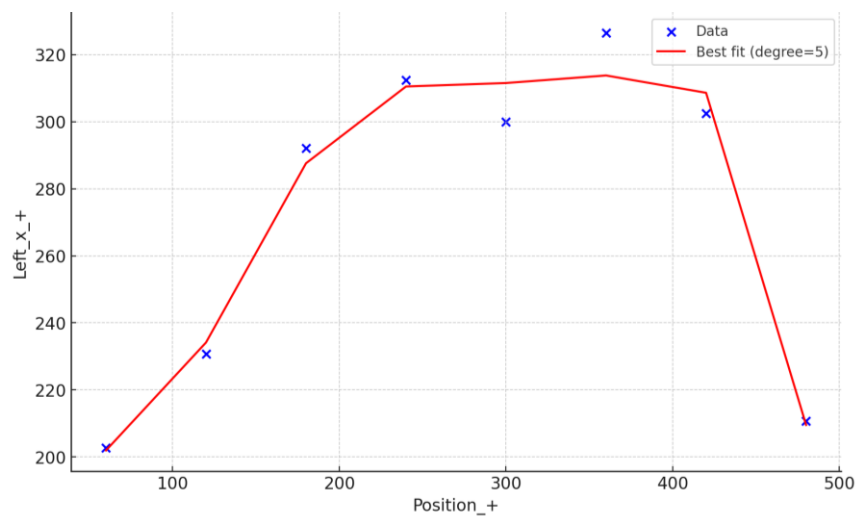
Díky prvnímu experimentu bylo zjištěno, jak ChatGPT data zpracovává a zda dokáže zpracovat regresní analýzu i z pohledu statistických významností jednotlivých koeficientů. Dále se testovalo, zda dospěje ke stejnému závěru, tedy regresní rovnici 2. stupně, která je uvedena v kapitole „9.2 Zpracování naměřených dat“. I když musel být jazykový model přímo naveden na regresní rovnici 2. stupně, tak lze tento experiment považovat za úspěch, protože položil základy prvnímu skutečnému promptu.

9.3.2 Druhý experiment – první skutečný prompt

První skutečný prompt byl vytvořen na základě poznatků získaných v rámci výše zmíněného experimentu. Cílem druhého experimentu bylo vložit požadavky do umělé inteligence tak, aby sama vyhodnotila, který regresní model je nejvýhodnější a zároveň byly všechny regresní koeficienty statisticky významné.

Vstup

Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde nezávisle proměnná bude position_+ a závisle proměnná je left_x_+. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé rovnice od 1. do 5. stupně a napiš nejlepší výsledek. + Position_+ left_x_+ 60 202,667 120 230,667 180 292,167 240 312,5 300 300 360 326,5 420 302,5 480 210,667 Proveď výpočet statistické významnosti parametrů regresní rovnice position_+ a dle nich uprav regresní rovnici. Statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05

Výstup

Obrázek 13. Zobrazení nejvhodnějšího regresního modelu dle ChatGPT [34]

Nejlepší regresní model je 5. stupně, s vysokou hodnotou $R^2 = 0,978$, což naznačuje velmi dobrou shodu modelu s daty. Přestože model vykazuje nízkou MSE (46,35), některé koeficienty regresní rovnice nejsou statisticky významné, protože jejich p-hodnoty jsou vyšší než 0,05.

Výsledky:

Koeficienty pro vyšší mocniny (x^3 , x^4 , x^5) mají p-hodnoty nad 0,05, což naznačuje, že nemusí významně přispívat k modelu.

Doporučuje se vytvořit nový model s odstraněním těchto nevýznamných koeficientů (redukovat stupeň polynomu) a znovu zhodnotit.

Pokud byste chtěl upravit model odstraněním nevýznamných koeficientů a provést další analýzu, mohu pokračovat. Jak si přejete postupovat?

Vstup

Ano, zkontroluj jednotlivé koeficienty a ukaž mi, které je možné redukovat

Výstup

Všechny koeficienty modelu mají p-hodnotu vyšší než 0,05, což znamená, že žádný z nich není statisticky významný na úrovni spolehlivosti 95 %. Nejbližší významnosti je konstanta (x_0), jejíž p-hodnota je 0,13.

Doporučuji:

Zredukovat stupeň polynomu a znovu provést test statistické významnosti.

Vyzkoušet regresní model nižšího stupně (například 4 nebo 3), protože vyšší stupně mohou vést k přefitování (overfitting) dat.

Mohu pokračovat v analýze s redukováným modelem (např. polynomem 4. stupně)?

Vstup

Ano

Výstup

Model s polynomem 4. stupně má $R^2 = 0,939$, což je mírně nižší než u modelu 5. stupně, ale stále velmi dobré. Nicméně žádný z koeficientů v tomto modelu není statisticky významný (p -hodnoty všech jsou vyšší než 0,05).

Doporučuji dále snížit stupeň polynomu na 3 nebo 2 a opět zkontrolovat, zda se statistická významnost zlepší. Chcete pokračovat s modelem 3. stupně?

Vstup

Ano

Výstup

Model s polynomem 3. stupně má $R^2 = 0,926$ a hodnoty koeficientů jsou blíže statistické významnosti než u vyšších stupňů. Konstanta (x_0) je statisticky významná ($p = 0,016$), ale ostatní koeficienty (x_1, x_2, x_3) mají stále $p > 0,05$, takže nejsou významné.

Další doporučení:

Zkusit ještě snížit stupeň polynomu na 2 a zjistit, zda se zvýší počet statisticky významných koeficientů.

Pokud ani model 2. stupně nebude mít více významných parametrů, zvážit návrat k jednodušší lineární regresi (1. stupeň).

Chcete pokračovat modelem 2. stupně?

Vstup

Ano

Výstup

Model s polynomem 2. stupně má $R^2 = 0,869$ a všechny jeho koeficienty jsou statisticky významné na úrovni spolehlivosti 95 %:

- Konstanta (x_0): $p = 0,013$
- Lineární člen (x_1): $p = 0,002$
- Kvadratický člen (x_2): $p = 0,003$

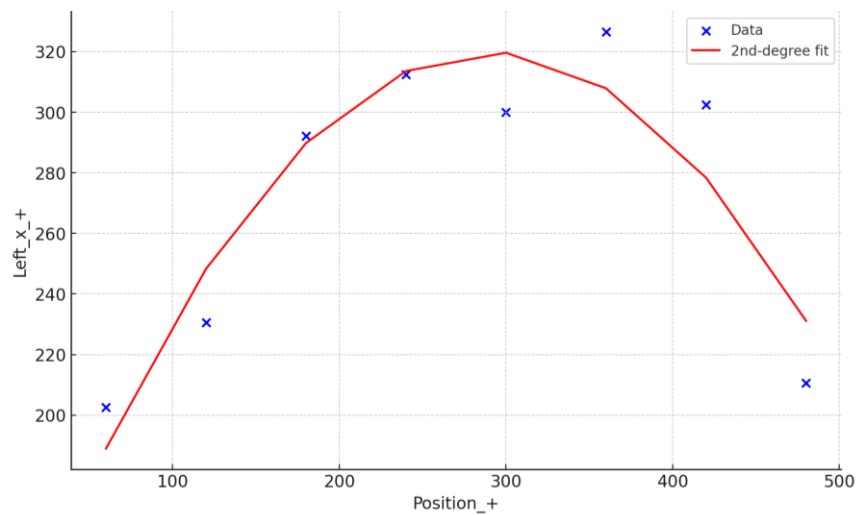
To znamená, že tento model je dobře vysvětlený s minimem přefitování, a všechny jeho parametry jsou důležité pro popis vztahu mezi proměnnými. Tento model tedy doporučuji jako finální řešení.

Pokud si přejete, mohu zobrazit jeho graf a rovnici, nebo vám poskytnout další analýzy. Jak chcete pokračovat?

Vstup

Zobraz graf a rovnici

Výstup



Obrázek 14. Výsledek prvního skutečného promptu [34]

Graf zobrazuje data a průběh regresní křivky 2. stupně. Rovnice regresního modelu je:

$$y = 111.905 + 1.432 \cdot x - 0.002467 \cdot x^2$$

Pokud potřebujete další výpočty nebo interpretace, dejte vědět!

Závěr druhého experimentu

I když ChatGPT zprvu označil jako nejlepší regresní model polynom 5. stupně, postupnými úpravami a kontrolou statistické významnosti jednotlivých regresních koeficientů dospěl ke stejnému správnému závěru, který je popsán v kapitole „9.2 Zpracování naměřených dat“. Tímto experimentem byly získány další poznatky, sloužící k dalšímu vylepšení vstupních parametrů tak, aby tyto vstupy mohl vkládat i uživatel bez hlubokých znalostí statistiky a získával statisticky správné výsledky i s vysvětlením.

9.3.3 Třetí experiment – rozdělení celého procesu

Cílem třetího experimentu je rozdělit celý proces na více jednodušších kroků, ke kterým bude uživatel postupně vyzván. Splněním jednotlivých dílčích operací obdrží výsledek s odpovídající regresní rovnicí.

Prvním krokem bylo vytvořit požadavek, který definuje, co má ChatGPT udělat a jakým způsobem má postupovat při vyhodnocení dat.

Nový vstupní požadavek

1. Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. Následně začni interpretovat výsledky od nejvyššího stupně (5. stupeň) a vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu. Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou.

Tento vstupní požadavek, v porovnání s předchozím, neobsahuje data. Jedná se pouze o zadání úkolu. Vložením věty „Pokud rozumíš, odpověz Vlož závisle proměnnou“ je zajištěno, že po zadání těchto instrukcí dojde pokaždé k vyzvání uživatele, aby vložil závisle proměnnou. Pokud by tato věta nebyla zadána, mohlo by se stát, že se jednotlivé reakce budou lišit a uživatel k tomuto kroku nebude vyzván.

V rámci třetího experimentu začíná konverzace výše zmíněným promptem.

Výstup

Rozumím! Vlož prosím závisle proměnnou (data Y). Jakmile je budu mít, požádám tě o nezávisle proměnnou (data X).

Vstup

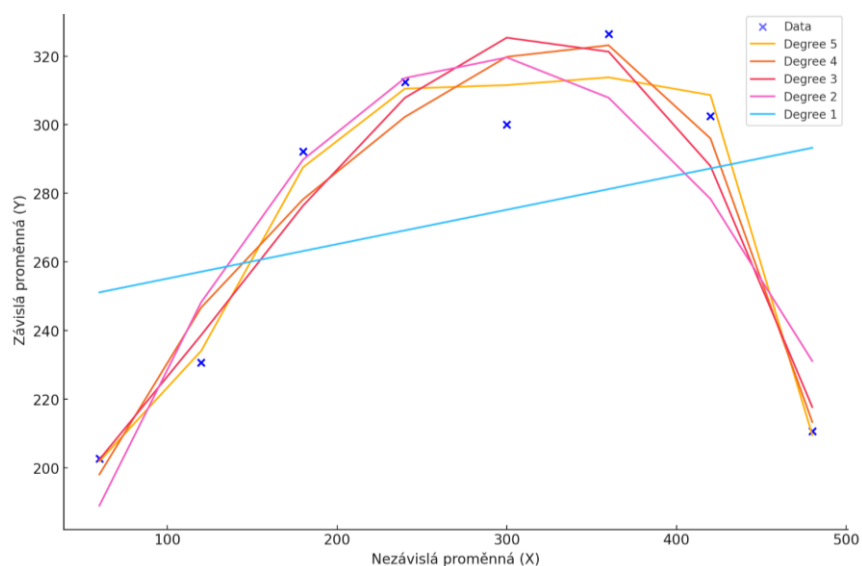
202,667 230,667 292,167 312,5 300 326,5 302,5 210,667

Výstup

Skvělé, máme závisle proměnnou (Y). Ted' prosím vlož data nezávislé proměnné (X).

Vstup

60 120 180 240 300 360 420 480

Výstup

Obrázek 15. Porovnání polynomických regresních modelů [34]

Tabulka 2. Výsledky polynomicke regrese [34]

<i>Stupeň</i>	<i>MSE</i>	<i>R²</i>	<i>P-hodnota</i>
5	46.349705532096706	0.9781990547633114	[0.13040165 0.31691705 0.24692161 0.22598923 0.21321117 0.199973]
4	128.76700926219092	0.93943343361106	[0.28128528 0.41293781 0.56923598 0.55470054 0.46783404]
3	158.275048944805	0.9255540971670155	[0.01619251 0.63832151 0.44764569 0.15495359]
2	279.5022989523808	0.8685339153067584	[0.01271969 0.0022948 0.00284132]
1	1935.79103705952	0.08948559857876726	[0.00081705]

Výsledky analýzy jednotlivých regresních modelů (od 1. do 5. stupně) jsou připraveny a zobrazeny v přehledu. Nyní mohu začít interpretací výsledků pro model 5. stupně. Chceš, abych pokračoval?

Vstup

Ano

Výstup

Model 5. stupně:

MSE (střední kvadratická chyba): 46.35

- Nejnižší MSE ze všech modelů, což znamená nejlepší přizpůsobení datům.

R^2 (koeficient determinace): 0.978

- Model vysvětluje 97,8 % variability v závislé proměnné, což je vynikající hodnota.

Statistická významnost parametrů (P-hodnoty):

- *P-hodnoty všech koeficientů jsou vyšší než 0,05, což znamená, že žádný z koeficientů není statisticky významný.*
- *Interpretace: Parametry modelu nelze na hladině významnosti 0,05 považovat za dostatečně spolehlivé.*

Model 5. stupně nabízí nejlepší přizpůsobení datům z hlediska MSE a R^2 . Nicméně absence statisticky významných parametrů naznačuje, že model je pravděpodobně přeinvestovaný (overfitted).

Mám pokračovat interpretací pro model 4. stupně?

Vstup

Ano

Výstup

Model 4. stupně:

MSE (střední kvadratická chyba): 128.77

- *Vyšší než u modelu 5. stupně, ale stále relativně nízká.*

R^2 (koeficient determinace): 0.939

- *Model vysvětluje 93,9 % variability v závislé proměnné, což je velmi dobré.*

Statistická významnost parametrů (P-hodnoty):

- *P-hodnoty všech koeficientů jsou vyšší než 0,05, což znamená, že žádný z parametrů není statisticky významný.*
- *Interpretace: Parametry modelu nelze na hladině významnosti 0,05 považovat za spolehlivé.*

Model 4. stupně vykazuje dobré přizpůsobení datům ($MSE = 128.77$, $R^2 = 0.939$), ale stejně jako model 5. stupně nemá žádné statisticky významné parametry. Tento model je méně přeinvestovaný než model 5. stupně, ale stále není ideální.

Mám pokračovat interpretací pro model 3. stupně?

Vstup

Ano

Výstup

Model 3. stupně:

MSE (střední kvadratická chyba): 158.28

- *Vyšší než u modelů 4. a 5. stupně, ale stále přijatelná.*

R² (koeficient determinace): 0.926

- *Model vysvětluje 92,6 % variability v závislé proměnné, což je dobrý výsledek.*

Statistická významnost parametrů (P-hodnoty):

- *Intercept: $P = 0.016$ – statisticky významný ($P < 0.05$).*
- *Koeficient X : $P = 0.638$ – není statisticky významný ($P > 0.05$).*
- *Koeficient X^2 : $P = 0.432$ – není statisticky významný ($P > 0.05$).*
- *Koeficient X^3 : $P = 0.298$ – není statisticky významný ($P > 0.05$).*

Model 3. stupně vykazuje solidní přizpůsobení datům ($MSE = 158.28$, $R^2 = 0.926$). Nicméně pouze intercept je statisticky významný ($P < 0.05$), zatímco všechny ostatní koeficienty nejsou. Tento model je jednodušší než modely 4. a 5. stupně, ale statistická významnost parametrů zůstává problematická.

Mám pokračovat interpretací pro model 2. stupně?

Vstup

Ano

Výstup

Model 2. stupně:

MSE (střední kvadratická chyba): 279.50

- *Výrazně vyšší než u modelů 3., 4. a 5. stupně, což znamená horší přizpůsobení datům.*

R² (koeficient determinace): 0.869

- *Model vysvětluje 86,9 % variability v závislé proměnné, což je přijatelná hodnota, ale nižší než u vyšších stupňů.*

Statistická významnost parametrů (P-hodnoty):

- *Intercept: $P = 0.013$ – statisticky významný ($P < 0.05$).*

- Koeficient X : $P = 0.002$ – statisticky významný ($P < 0.05$).
- Koeficient X^2 : $P = 0.013$ – statisticky významný ($P < 0.05$).

Model 2. stupně vykazuje horší přizpůsobení datům ($MSE = 279.50$, $R^2 = 0.869$) než vyšší modely, ale jeho parametry (intercept, X a X^2) jsou všechny statisticky významné ($P < 0.05$). Tento model je jednodušší a zároveň obsahuje důvěryhodné koeficienty, což z něj činí kandidáta na nejvhodnější model.

Mám pokračovat interpretací pro model 1. stupně?

Vstup

Ano

Výstup

Model 1. stupně (lineární regrese):

MSE (střední kvadratická chyba): 1935.79

- Nejvyšší hodnota mezi všemi modely, což znamená nejhorší přizpůsobení datům.

R^2 (koeficient determinace): 0.089

- Model vysvětluje pouze 8,9 % variability v závislé proměnné, což je velmi nízké.

Statistická významnost parametrů (P -hodnoty):

- Intercept: $P = 0.001$ – statisticky významný ($P < 0.05$).
- Koeficient X : $P = 0.472$ – není statisticky významný ($P > 0.05$).

Model 1. stupně (lineární regrese) má nejhorší přizpůsobení datům (nejvyšší MSE a velmi nízké R^2). I když je intercept statisticky významný, samotný koeficient XXX není. Tento model nelze považovat za vhodný.

Celkové doporučení:

Model 2. stupně je nejvhodnější, protože:

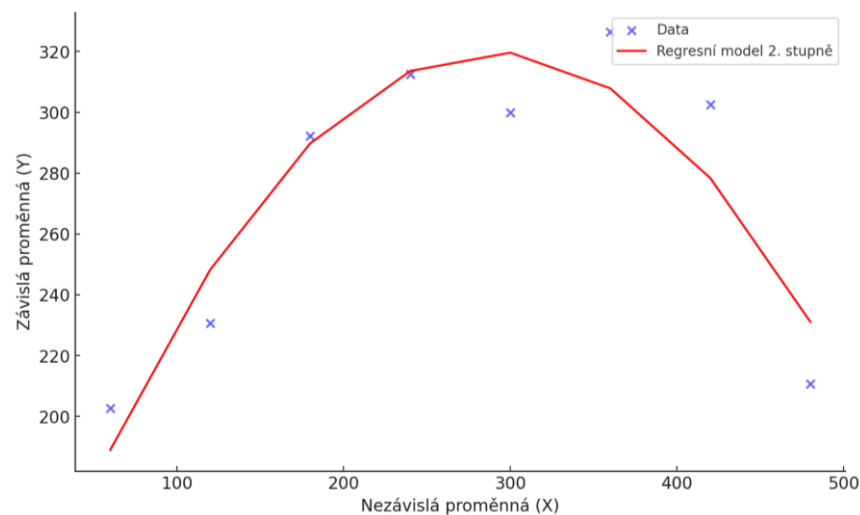
- Všechny jeho parametry jsou statisticky významné ($P < 0.05$).
- Přizpůsobení datům ($MSE = 279.50$, $R^2 = 0.869$) je dobré, i když horší než u modelů vyšších stupňů.
- Model je dostatečně jednoduchý, aby nebyl přeinvestovaný.

Pokud chceš, mohu poskytnout přesnou regresní rovnici modelu 2. stupně, nebo další vizualizaci.

Vstup

Poskytni

Výstup



Obrázek 16. Regresní model 2. stupně [34]

Regresní rovnice modelu 2. stupně je následující:

$$y = 111.905 + 1.432x - 0.002x^2$$

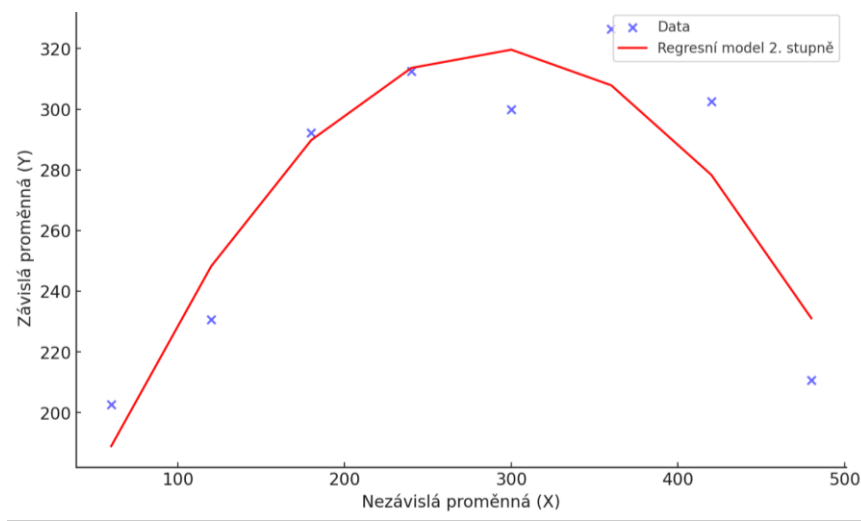
Graf výše zobrazuje data (modré body) a odpovídající model 2. stupně (červená čára).

Pokud chceš provést další analýzu nebo vizualizaci, dej vědět!

Vstup

Poskytni přesné rovnice pro nejvýhodnější regresní rovnici, vizualizuj graf a vypiš všechny potřebné parametry: korelační koeficient, koeficient determinace, střední kvadratickou chybu MEP a akaikeho informační kritérium.

Výstup



Obrázek 17. Regresní model 2. stupně – nejvýhodnější reg. Rovnice [34]

Parametry nejvýhodnější regresní rovnice (model 2. stupně):

1. Regresní rovnice:

- $y = 111.905 + 1.432x - 0.002x^2$

2. Korelační koeficient: 0.932

- Silná pozitivní korelace mezi skutečnými hodnotami a predikcí.

3. Koeficient determinace (R^2): 0.869

- Model vysvětluje 86,9 % variability závislé proměnné.

4. Střední kvadratická chyba (MSE): 279.50

- Průměrná velikost chyby modelu.

5. Akaikeho informační kritérium (AIC): 53.06

- Penalizuje složitost modelu; nižší hodnota je lepší.

Graf výše zobrazuje regresní model 2. stupně (červená čára) spolu s daty (modré body).

Pokud potřebuješ další informace nebo interpretaci, dej vědět!

Úprava posledního vstupu

Poskytni přesné rovnice pro nejvýhodnější regresní rovnici, vizualizuj graf a vypiš všechny potřebné parametry: korelační koeficient, koeficient determinace, střední kvadratickou chybu predikce MEP a akaikeho informační kritérium **na tři desetinná místa**.

Byl přidán červeně zvýrazněný dodatek, aby byly veškeré výstupní hodnoty zaokrouhleny na tři desetinná místa.

Výstup po úpravě

Parametry nejvýhodnější regresní rovnice (model 2. stupně) s přesností na tři desetinná místa:

1. Regresní rovnice:

- $y = 111.905 + 1.432x - 0.002x^2$

2. Korelační koeficient: 0.932

- *Silná pozitivní korelace mezi skutečnými hodnotami a predikcí.*

3. Koeficient determinace (R^2): 0.869

- *Model vysvětluje 86,9 % variability závislé proměnné.*

4. Střední kvadratická chyba (MSE): 279.502

- *Odhad průměrné velikosti kvadratické chyby modelu.*

5. Průměrná chyba predikce (MEP): 14.695

- *Průměrná absolutní chyba mezi predikovanými a skutečnými hodnotami.*

6. Akaikeho informační kritérium (AIC): 53.064

- *Hodnota penalizující složitost modelu; nižší je lepší.*

Graf výše zobrazuje model 2. stupně (červená čára) spolu s daty (modré body). Pokud máš zájem o další rozšíření nebo úpravy, dej vědět!

Vstup

Zapomeň vše, v rámci této konverzace

Výstup

Informace uložená v paměti byla aktualizována

Hotovo, všechny informace z této konverzace byly zapomenuty. Pokud budeš něco potřebovat, dej vědět!

Závěr třetího experimentu

V rámci tohoto experimentu vznikl první skutečný prompt obsahující osm základních textů, které bude uživatel zadávat a ChatGPT odpoví téměř totožně, jako ve výše zmíněných výstupech. Aktuální vypadá následovně:

1. Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. Následně začni interpretovat výsledky od nejvyššího stupně (5. stupeň) a vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu. Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou.

2. Uživatel vloží závisle proměnnou a následně nezávisle proměnnou

3. Ano – chce začít interpretovat výsledky pro model 5. stupně

4. Ano – pokračovat na interpretaci modelu 4. stupně

5. Ano – pokračovat na interpretaci modelu 3. stupně

6. Ano – pokračovat na interpretaci modelu 2. stupně

7. Ano – pokračovat na interpretaci modelu 1. stupně

8. Poskytni – uživatel chce poskytnout přesnou rovnici nejvýhodnějšího regresního modelu

9. Poskytni přesné rovnice pro nejvýhodnější regresní rovnici, vizualizuj graf a vypiš všechny potřebné parametry: korelační koeficient, koeficient determinace, střední kvadratickou chybu predikce MEP a akaikého informační kritérium na tři desetinná místa.

Zadáním posledního promptu obdrží uživatel stejný výsledek, jaký je popsán výše v rámci třetího experimentu.

10. Zapomeň vše, v rámci této konverzace – ChatGPT vymaže mezipaměť a může dojít k opětovnému zadávání jednotlivých promptů a vyhodnocení jiných dat.

Třetí experiment přinesl velmi cenné informace související se zadáváním jednotlivých příkazů a jejich výstupy. Původní text byl rozdělen na více částí, přičemž uživatel jako první

zadáva požadavek a následně vkládá jednotlivá data, která jsou postupně vyhodnocena. Na závěr je interpretován výsledek a je vysvětlen význam jednotlivých parametrů.

Aktuálním nedostatkem jsou neošetřená vstupní data. Zatím není definováno, co se má stát, pokud by uživatel vložil neplatné hodnoty (text místo čísel) nebo jiný počet hodnot. Zároveň si vložená data nelze prohlédnout a zkontrolovat jejich správnost.

9.3.4 Čtvrtý experiment – ošetření vstupů a ladění textů

Nyní je nezbytné zabezpečit, aby v případě špatně zadaných vstupních dat nedocházelo ke zpracování regresní analýzy, ale namísto toho byl uživatel upozorněn na chybu a ChatGPT žádal opravu. Za chybu jsou považována textová data, či neshodný počet hodnot závislé a nezávislé proměnné. První prompt, který tyto informace obsahuje je zmíněn níže.

První prompt zaměřený na kontrolu vstupu

1. Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. Následně začni interpretovat výsledky od nejvyššího stupně (5. stupeň) a vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu. Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou. Zároveň chtěj, aby byla vložena data numerického typu. Pokud se bude jednat o text, žádej od uživatele opravu. Také zkontroluj počet vložených dat u závislé a nezávislé proměnné. Vytvoř jednoduchou tabulku, kde bude závislá a nezávislá proměnná a v případě nedodržení číselného formátu uživatele, či jiného počtu hodnot upozorni na chybu.

Výstup

Rozumím! Vlož prosím závisle proměnnou (data Y).

- *Ujisti se, že vkládáš pouze číselné hodnoty (bez textu).*
- *Jakmile je budu mít, požádám tě o nezávisle proměnnou (data X).*
- *Počet hodnot ve vstupních datech musí být stejný pro obě proměnné, jinak tě na to upozorním.*

Čekám na vstup dat.

Vstup

202,667 230,667 292,167 312,5 300 326,5 302,5

Výstup

Zdá se, že jsi vložil 7 hodnot pro závislou proměnnou.

Prosím, potvrď, zda je to správně a zda chceš pokračovat s těmito daty. Pokud jsi chtěl vložit jiný počet hodnot, doplň chybějící nebo uprav formát tak, aby všechny hodnoty byly numerické.

Jakmile potvrdíš správnost, požádám tě o nezávisle proměnnou (data X).

Vstup

Ano

Výstup

Děkuji za potvrzení! Nyní prosím vlož nezávisle proměnnou (data X).

- *Musí obsahovat stejný počet hodnot jako závislá proměnná (7 hodnot).*
- *Ujistí se, že jsou všechny hodnoty číselné.*

Čekám na vstup dat.

Vstup

60 120 180 240 300 360 420 480

Výstup

Zdá se, že jsi vložil 8 hodnot pro nezávislou proměnnou, zatímco pro závislou proměnnou jsi vložil 7 hodnot.

Abych mohl správně provést regresní analýzu, je nutné, aby obě proměnné měly stejný počet hodnot.

Prosím, oprav počet hodnot tak, aby odpovídal, a znovu vlož nezávisle proměnnou (data X).

Konec experimentu

Tímto výstupem experiment končil, protože ChatGPT neporozuměl správně zadání a nedodržel všechny body, které byly zmíněny. V zadání je zmíněna věta: „Vytvoř jednoduchou tabulku, kde bude závislá a nezávislá proměnná a v případě nedodržení číselného formátu uživatele, či jiného počtu hodnot upozorni na chybu.“ ChatGPT žádnou tabulku nevytvořil. Z uvedeného důvodu je nutné prompt přepsat a znovu otestovat.

Upravená verze vstupního promptu

1. Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. Následně začni interpretovat výsledky od nejvyššího stupně (5. stupeň) a vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu.

~~Zároveň chtěj, aby byla vložena data numerického typu.~~ **Zároveň zkontroluj, aby byla vložena data numerického typu.** Pokud se bude jednat o text, žádej od uživatele opravu. ~~Také zkontroluj počet vložených dat u závislé a nezávislé proměnné.~~ **Také sám pro sebe zkontroluj počet vložených hodnot u obou proměnných.** ~~Vytvoř jednoduchou tabulku, kde bude závislá a nezávislá proměnná a v případě nedodržení číselného formátu uživatele, či jiného počtu hodnot upozorni na chybu.~~ **Ihned po vložení závisle proměnné vytvoř tabulku, která bude obsahovat dva sloupce. První sloupec – číslování hodnot, druhý sloupec hodnoty závisle proměnné. Jakmile tuto tabulku vytvoříš, žádej uživatele o vložení nezávisle proměnné. Jakmile uživatel vloží, doplň k tabulce třetí sloupec, který pojmenuj nezávisle proměnná. Pokud se závislá a nezávislá proměnná bude lišit, chtěj po uživateli opětovné zadání obou proměnných. V případě jakéhokoliv nesouladu upozorni uživatele na chybu. Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou.**

Tučně a červeně zvýrazněný text označuje nové úpravy, které byly v promptu provedeny. Některé věty byly pouze parafrázovány a doplněny o další informace.

Výstup

Vlož závisle proměnnou.

Vstup

202,667 230,667 292,167 312,5 300 326,5 302,5 210,667

Výstup

Tabulka 3. Chybné zobrazení hodnot [34]

<i>Číslo</i>	<i>Závislá proměnná</i>	<i>Nezávislá proměnná</i>
1	202667	
2	230667	
3	292167	
4	312500	
5	300000	
6	326500	
7	302500	
8	210667	

Zhodnocení výstupu

Díky nově definovanému promptu vytvořil ChatGPT tabulku, ale lze vidět, že data neodpovídají vloženým hodnotám (neobsahují desetinnou čárku). Z tohoto důvodu bylo nezbytné dále definovat, jak mají být data zobrazena a přidat informaci týkající se desetinných čárek a míst.

Upravený vstup – ošetření desetinných čárek a míst

1. Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. Následně začni interpretovat výsledky od nejvyššího stupně (5. stupeň) a vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu.

Zároveň zkontroluj, aby byla vložena data numerického typu. **K oddělení desetinných míst jsou v datech využity čárky. Mezera odděluje jednotlivé hodnoty.** Pokud se bude jednat o text, žádej od uživatele opravu. Také sám pro sebe zkontroluj počet vložených hodnot u obou proměnných. Ihned po vložení závislé proměnné vytvoř tabulku, která bude

obsahovat dva sloupce. **Ke každé z hodnot v tabulce zobraz tři desetinná místa, i když se bude jednat o celé číslo.** První sloupec – číslování hodnot, druhý sloupec hodnoty závisle proměnné. Jakmile tuto tabulku vytvoříš, žádej uživatele o vložení nezávisle proměnné. Jakmile uživatel vloží, doplň k tabulce třetí sloupec, který pojmenuj nezávisle proměnná. Pokud se závislá a nezávislá proměnná bude lišit, vlož hodnoty do tabulky a chtěj po uživateli opětovné zadání obou proměnných. V případě jakéhokoliv nesouladu upozorni uživatele na chybu. Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou.

Výstup po vložení hodnot závisle proměnné

Tabulka 4. Správný formát vložených hodnot [34]

<i>Číslo</i>	<i>Závislá proměnná</i>	<i>Nezávislá proměnná</i>
1	202.667	
2	230.667	
3	292.167	
4	312.500	
5	300.000	
6	326.500	
7	302.500	
8	210.667	

Z tabulky lze vyčíst, že bližším definováním vstupních hodnot ChatGPT upravil data v tabulce. Následně konverzace pokračovala vložím hodnot nezávislé proměnné.

Výstup po vložení hodnot nezávisle proměnné

Počet vložených hodnot nezávislé proměnné (7 hodnot) se neshoduje s počtem hodnot závislé proměnné (8 hodnot).

Prosím, oprav data a zadej obě proměnné znovu, aby měly stejný počet hodnot.

V tomto okamžiku byl experiment přerušen a došlo k opětovnému opravení vstupního promptu, protože původním záměrem bylo, aby ChatGPT zobrazil hodnoty nezávislé proměnné v tabulce a následně začal s regresní analýzou.

Upravení vstupu – zobrazení hodnot nezávisle proměnné v tabulce

1. Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. Následně začni interpretovat výsledky od nejvyššího stupně (5. stupeň) a vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu.

Zároveň zkontroluj, aby byla vložena data numerického typu. K oddělení desetinných míst jsou v datech využity čárky. Mezera odděluje jednotlivé hodnoty. Pokud se bude jednat o text, žádej od uživatele opravu. Také sám pro sebe zkontroluj počet vložených hodnot u obou proměnných. Ihned po vložení závisle proměnné vytvoř tabulku, která bude obsahovat dva sloupce. Ke každé z hodnot v tabulce zobraz tři desetinná místa, i když se bude jednat o celé číslo. První sloupec – číslování hodnot, druhý sloupec hodnoty závisle proměnné. Jakmile tuto tabulku vytvoříš, žádej uživatele o vložení nezávisle proměnné. Jakmile uživatel vloží, doplň k tabulce třetí sloupec, který pojmenuj nezávisle proměnná. Pokud se závislá a nezávislá proměnná bude lišit, ~~vlož hodnoty do tabulky a chtěj po uživateli opětovné zadání obou proměnných.~~ **postupuj následovně. Aktualizuj tabulku hodnot – doplň vložené hodnoty nezávisle proměnné. Po zobrazení hodnot uživateli zmiň, tyto skutečnosti a žádej po opětovné zadání závisle proměnné a opět aktualizuj tabulku. Po aktualizaci vyzvi k opravnému vložení nezávisle proměnné a opětovně aktualizuj tabulku.** V případě jakéhokoliv nesouladu upozorni uživatele na chybu. Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou.

Další testování promptu

V rámci dalšího testování byly opět vloženy hodnoty závislé a nezávislé proměnné. Nyní bylo cílem zjistit, jak se ChatGPT zachová, pokud by uživatel vložil jiný počet hodnot závislé proměnné a nezávislé proměnné. Z výše zmíněného důvodu bylo v rámci závislé proměnné vloženo 8 hodnot a v rámci nezávislé proměnné 7 hodnot. Výstup ChatGPT je zachycen níže.

Výstup ChatGPT při vložení rozdílného počtu hodnot

Tabulka 5. Výsledek rozdílného počtu hodnot [34]

<i>Číslo</i>	<i>Závislá proměnná</i>	<i>Nezávislá proměnná</i>
<i>1</i>	<i>202.667</i>	<i>120.000</i>
<i>2</i>	<i>230.667</i>	<i>180.000</i>
<i>3</i>	<i>292.167</i>	<i>240.000</i>
<i>4</i>	<i>312.500</i>	<i>300.000</i>
<i>5</i>	<i>300.000</i>	<i>360.000</i>
<i>6</i>	<i>326.500</i>	<i>420.000</i>
<i>7</i>	<i>302.500</i>	<i>480.000</i>
<i>8</i>	<i>210.667</i>	

Počet hodnot nezávislé proměnné (7 hodnot) nesouhlasí s počtem hodnot závislé proměnné (8 hodnot).

Níže je aktualizovaná tabulka s vloženými hodnotami:

Aktualizovaná tabulka s hodnotami:

Závislá proměnná:

- *202.667*
- *230.667*
- *292.167*
- *312.500*
- *300.000*
- *326.500*
- *302.500*
- *210.667*

Nezávislá proměnná:

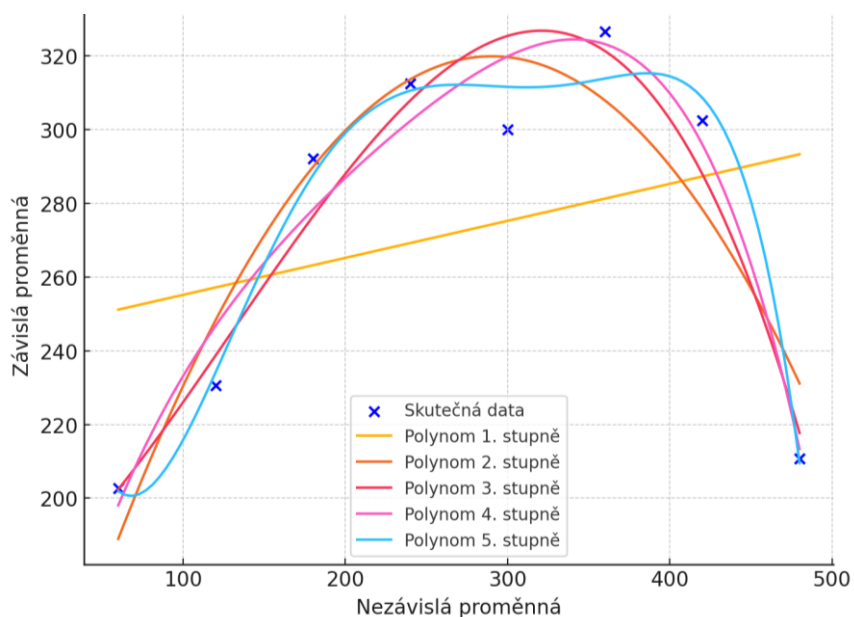
- 120
- 180
- 240
- 300
- 360
- 420
- 480
- (chybí jedna hodnota)

Prosím, vložená data zkontroluj a zadej znovu závislou proměnnou. Poté budeš vyzván k opravnému vložení nezávislé proměnné.

Po opětovném zadání správného počtu hodnot dochází k opětovnému zobrazení tabulky a textovému výstupu:

Data byla správně vložena. Nyní mohu pokračovat v analýze.

Začnu porovnáním regresních rovnic od 1. do 5. stupně a výpočtem statistické významnosti parametrů. Pustím se do toho.



Obrázek 18. Regresní modely různých stupňů [34]

Výsledky pro polynom 5. stupně ukazují $R^2 = 0.978$, což znamená velmi dobré přizpůsobení modelu.

Nicméně, žádný z koeficientů není statisticky významný (p -hodnoty všech koeficientů jsou vyšší než 0.05).

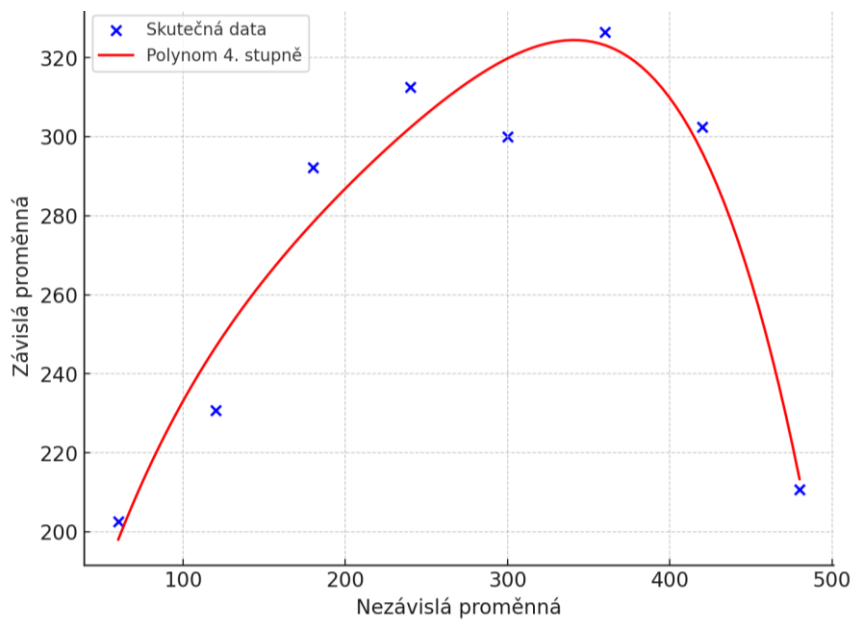
To znamená, že model může být přeucený a některé z vyšších polynomických členů nemusí být potřebné.

Mám pokračovat k analýze polynomu 4. stupně? (ano/ne)

Vstup

Ano

Výstup



Obrázek 19. Regresní model – Polynom 4. stupně [34]

Výsledky pro polynom 4. stupně ukazují $R^2 = 0.939$, což je stále velmi dobré přizpůsobení modelu.

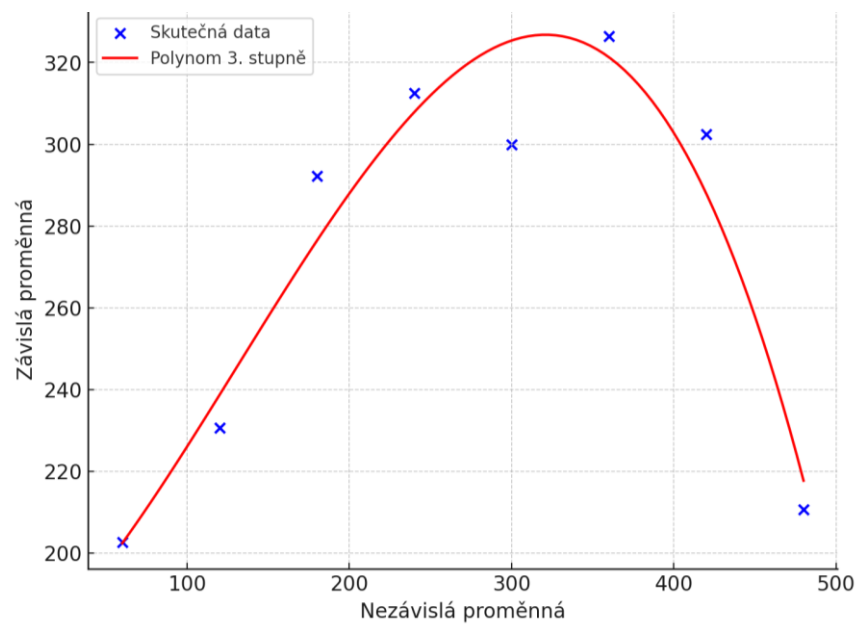
Nicméně, žádný z koeficientů není statisticky významný (p -hodnoty všech koeficientů jsou vyšší než 0.05).

To naznačuje, že tento model obsahuje nepotřebné vyšší mocniny nezávislé proměnné.

Mám pokračovat k analýze polynomu 3. stupně? (ano/ne)

Vstup

Ano

Výstup

Obrázek 20. Regresní model – Polynom 3. stupně [34]

Výsledky pro polynom 3. stupně ukazují $R^2 = 0.926$, což je stále velmi dobré přizpůsobení modelu.

Statistická významnost koeficientů:

- Pouze konstantní člen (intercept) je statisticky významný ($p = 0.016$)
- Ostatní koeficienty nejsou významné ($p > 0.05$)

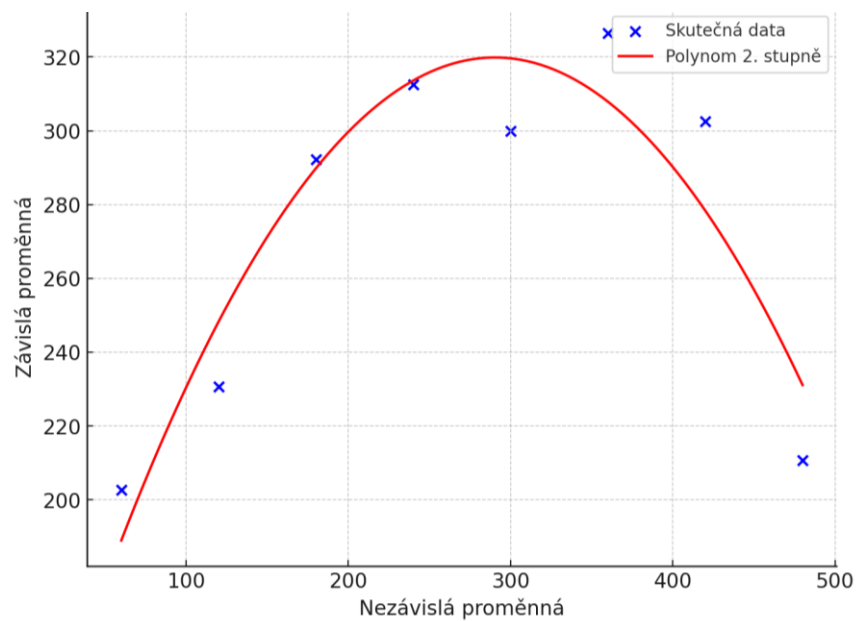
To naznačuje, že třetí stupeň polynomu nemusí být nutný a model může být zjednodušen.

Mám pokračovat k analýze polynomu 2. stupně? (ano/ne)

Vstup

Ano

Výstup



Obrázek 21. Regresní model – Polynom 2. stupně [34]

Výsledky pro polynom 2. stupně ukazují $R^2 = 0.869$, což je stále dobré přizpůsobení modelu.

Statistická významnost koeficientů:

- Všechny koeficienty jsou statisticky významné ($p < 0.05$)
- Model je tedy dobře vysvětlen kvadratickou rovnicí

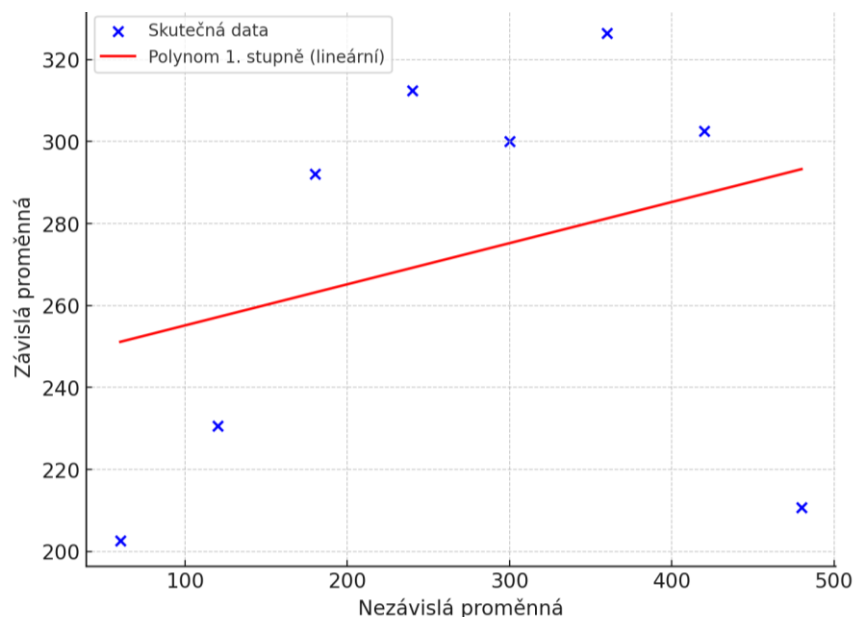
To naznačuje, že druhý stupeň polynomu je dostatečný a není nutné používat vyšší polynomy.

Mám pokračovat k analýze polynomu 1. stupně? (ano/ne)

Vstup

Ano

Výstup



Obrázek 22. Regresní model – Polynom 1. stupně [34]

Výsledky pro polynom 1. stupně (lineární regrese)

- $R^2 = 0.089$, což znamená velmi slabé přizpůsobení modelu
- Pouze konstantní člen (intercept) je statisticky významný
- Koeficient u nezávislé proměnné je nevýznamný ($p = 0.472 > 0.05$)

Shrnutí analýzy:

- Polynom 1. stupně (lineární regrese) není vhodný, protože nevysvětluje dobře závislost mezi proměnnými.
- Polynom 2. stupně je nejvhodnější model, protože všechny jeho koeficienty jsou statisticky významné a dobře popisují data.

Pokud chceš, mohu vypsát konkrétní regresní rovnici pro polynom 2. stupně. Chceš ji zobrazit? (ano/ne)

Vstup

Ano

Výstup

Nejlepší regresní rovnice (polynom 2. stupně):

$$y = -0.00247x^2 + 1.43247x + 111.90543$$

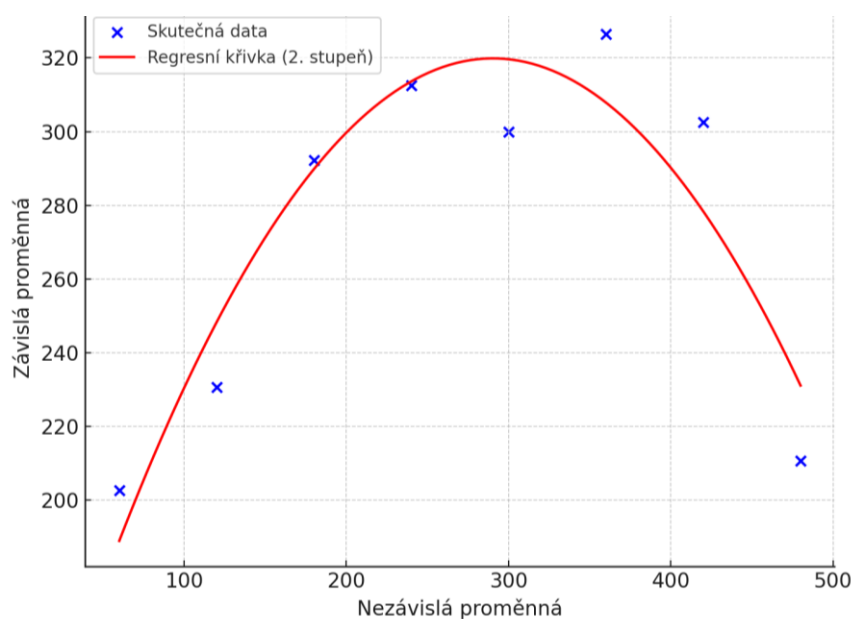
Tato rovnice nejlépe popisuje vztah mezi závislou a nezávislou proměnnou.

Mohu ti s něčím dalším pomoci?

Vstup

Poskytni přesné rovnice pro nejvýhodnější regresní rovnici, vizualizuj graf a vypiš všechny potřebné parametry: korelační koeficient, koeficient determinace, střední kvadratickou chybu predikce MEP a akaikeho informační kritérium na tři desetinná místa

Výstup



Obrázek 23. Nejlepší regresní model – Polynom 2. stupně [34]

Nejlepší regresní rovnice (polynom 2. stupně):

$$y = -0.00247x^2 + 1.43247x + 111.90543$$

Další klíčové parametry modelu:

- *Korelační koeficient (R): 0.932*
- *Koeficient determinace (R^2): 0.869*
- *Střední kvadratická chyba predikce (MEP): 16.718*
- *Akaikeho informační kritérium (AIC): 73.767*

Na grafu je vidět, že polynom 2. stupně dobře vystihuje vztah mezi proměnnými.

Mohu pro tebe udělat ještě něco dalšího?

Vyhodnocení posledních změn

Díky výše uvedeným změnám se veškeré hodnoty závislé i nezávislé proměnné vkládají do tabulky a zároveň ChatGPT hlídá správnost zadaných dat. Bohužel se kvůli těmto úpravám změnil výstupní informace poskytované během vyhodnocování jednotlivých polynomů. ChatGPT již nadále nezobrazuje jednotlivé koeficienty X a jejich statistické významnosti. V dalších úpravách proto bude kladen důraz na to, jakým způsobem mají být jednotlivé stupně regresních modelů analyzovány.

Nově upravený prompt

1. Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. ~~Následně začni interpretovat výsledky od nejvyššího stupně (5. stupeň) a vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu.~~

Následně porovnej všechny regresní modely. Využij k tomu tabulku a zobraz všechny regresní modely do jednoho grafu. Po tomto kroku se zeptej uživatele, zda můžeš začít interpretovat výsledky od nejvyššího stupně (5. stupeň). Vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu. V rámci interpretace (pouze u nejvyššího stupně) seznam uživatele s tím hodnotou MSE a tím, co představuje, tak aby tomu rozuměl i zcela neznalý člověk, koeficientem determinace a statistickou významností všech parametrů. Zobraz všechny parametry a jejich hodnoty. U dalších modelů jen zobraz

jednotlivé hodnoty MSE, koeficientu determinace a významnosti jednotlivých parametrů dle stupně regresního modelu.

Zároveň zkontroluj, aby byla vložena data numerického typu. K oddělení desetinných míst jsou v datech využity čárky. Mezera odděluje jednotlivé hodnoty. Pokud se bude jednat o text, žádej od uživatele opravu. Také sám pro sebe zkontroluj počet vložených hodnot u obou proměnných. Ihned po vložení závisle proměnné vytvoř tabulku, která bude obsahovat dva sloupce. Ke každé z hodnot v tabulce zobraz tři desetinná místa, i když se bude jednat o celé číslo. První sloupec – číslování hodnot, druhý sloupec hodnoty závisle proměnné. Jakmile tuto tabulku vytvoříš, žádej uživatele o vložení nezávisle proměnné. Jakmile uživatel vloží, doplň k tabulce třetí sloupec, který pojmenuj nezávisle proměnná. Pokud se závislá a nezávislá proměnná bude lišit, postupuj následovně. Aktualizuj tabulku hodnot – doplň vložené hodnoty nezávisle proměnné. Po zobrazení hodnot uživateli zmiň, tyto skutečnosti a žádej po opětovné zadání závisle proměnné a opět aktualizuj tabulku. Po aktualizaci vyzvi k oprávnému vložení nezávisle proměnné a opětovně aktualizuj tabulku. V případě jakéhokoli nesouladu upozorni uživatele na chybu. **Ještě předtím, než začneš vyhodnocovat data, chtěj potvrzení uživatelem.** Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou.

9.3.5 První chyba během analýzy dat

V porovnání s původním jednoduchým a krátkým promptem je aktuální prompt velmi obsáhlý a dlouhý. Při testování těchto změn došlo k prvnímu většímu problému. ChatGPT zobrazil vložené hodnoty v tabulkách, ale při interpretaci výsledků nezobrazil vůbec nic, pouze se dotázal, zda je možné pokračovat k analýze polynomu nižšího stupně. Chyba byla s největší pravděpodobností způsobena délkou promptu. Kvůli rozsahu nebyl ChatGPT schopen všechny informace pojmout a zpracovat. Výsledkem byla pouze věta odkazující na analýzu polynomu nižšího. Touto chybou se jasně definoval další nezbytný krok k vytvoření automatického systému pro nalezení nejvhodnější regrese. Původní příkaz je potřeba rozdělit na více kratších částí, které uživatel postupně vloží a až následně zadá hodnoty.

Mohu začít interpretovat výsledky od nejvyššího stupně (5. stupeň)? (ano/ne) [↩]

Mohu pokračovat k analýze polynomu 4. stupně? (ano/ne) [↩]

Mohu pokračovat k analýze polynomu 3. stupně? (ano/ne) [↩]

Please continue writing.

Zdá se, že došlo k resetování systému. Mohu začít znovu s analýzou dat, nebo chceš pokračovat v předchozím úkolu? 😊 [↩]

9. Zruš všechny požadované příkazy a zapomeň na vše v rámci této konverzace

Obrázek 24. Chyba způsobená příliš dlouhým zadáním promptu [34]

9.3.6 Finalizace promptu

V rámci finalizace došlo k rozdělení předchozího – velmi dlouhého promptu na několik kratších úseků, ve kterých jsou jasně definovány příkazy, které má ChatGPT provést. První text obsahuje požadavek na provedení regresní analýzy dat a způsob, jakým bude ChatGPT vložená data zobrazovat tzn. vkládání do tabulky. Přesný způsob a úpravu promptu, aby byla data vložena do tabulky, lze nalézt v kapitole 9.3.4 Čtvrtý experiment – ošetření vstupů a ladění textů. Nově rozdělený text čerpá ze všech poznatků získaných v rámci předešlých experimentů.

1. Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. **Jako první zkontroluj, aby byla vložena data numerického typu.** ~~Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. Následně porovnej všechny regresní modely. Využij k tomu tabulku a zobraz všechny regresní modely do jednoho grafu. Po tomto kroku se zeptej uživatele, zda můžeš začít interpretovat výsledky od nejvyššího stupně (5. stupeň). Vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu. V rámci interpretace (pouze u nejvyššího stupně) seznam~~

~~uživatele s tím hodnotou MSE a tím, co představuje, tak aby tomu rozuměl i zcela neznalý člověk, koeficientem determinace a statistickou významností všech parametrů. Zobraz všechny parametry a jejich hodnoty. U dalších modelů jen zobraz jednotlivé hodnoty MSE, koeficientu determinace a významnosti jednotlivých parametrů dle stupně regresního modelu.~~

~~Zároveň zkontroluj, aby byla vložena data numerického typu. K oddělení desetinných míst jsou v datech využity čárky. Mezera odděluje jednotlivé hodnoty. Pokud se bude jednat o text, žádej od uživatele opravu. Také sám pro sebe zkontroluj počet vložených hodnot u obou proměnných. Ihned po vložení závisle proměnné vytvoř tabulku, která bude obsahovat dva sloupce. Ke každé z hodnot v tabulce zobraz tři desetinná místa, i když se bude jednat o celé číslo. První sloupec – číslování hodnot, druhý sloupec hodnoty závisle proměnné. Jakmile tuto tabulku vytvoříš, žádej uživatele o vložení nezávisle proměnné. Jakmile uživatel vloží, doplň k tabulce třetí sloupec, který pojmenuj nezávisle proměnná. Pokud se závislá a nezávislá proměnná bude lišit, postupuj následovně. Aktualizuj tabulku hodnot – doplň vložené hodnoty nezávisle proměnné. Po zobrazení hodnot uživateli zmiň, tyto skutečnosti a žádej po opětovné zadání závisle proměnné a opět aktualizuj tabulku. Po aktualizaci vyzvi k opravnému vložení nezávisle proměnné a opětovně aktualizuj tabulku. V případě jakéhokoli nesouladu upozorni uživatele na chybu. Ještě předtím, než začneš vyhodnocovat data, chtěj potvrzení uživatelem. Pokud rozumíš, odpověz Vlož **další instrukce**. závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou.~~

Tímto textem by mělo dojít k ošetření vstupních hodnot. Poslední věta zajišťuje, že ChatGPT nebude pracovat na analýze dat, ale dovolí uživateli nahrát další instrukce. Konkrétně pouze zopakuje větu „Vlož další instrukce“.

2. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. Následně porovnej všechny regresní modely. Využij k tomu tabulku a zobraz všechny regresní modely do jednoho grafu. Po tomto kroku se zeptej uživatele, zda můžeš začít interpretovat výsledky od nejvyššího stupně (5. stupeň). Vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu. V rámci interpretace (pouze u nejvyššího stupně) seznam uživatele s tím hodnotou MSE a tím, co představuje, tak aby tomu rozuměl i zcela neznalý člověk, koeficientem determinace a statistickou významností

všech parametrů. **Výsledky interpretace zobrazuj vždy celé, bez možnosti jejich sbalení.** Zobraz všechny parametry a jejich hodnoty. U dalších modelů jen zobraz jednotlivé hodnoty MSE, koeficientu determinace a významnosti jednotlivých parametrů dle stupně regresního modelu. **Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěl i nezávisle proměnnou.**

Výstup po vložení hodnot závisle a nezávisle proměnné

Tabulka 6. Vložení závislé a nezávislé proměnné [34]

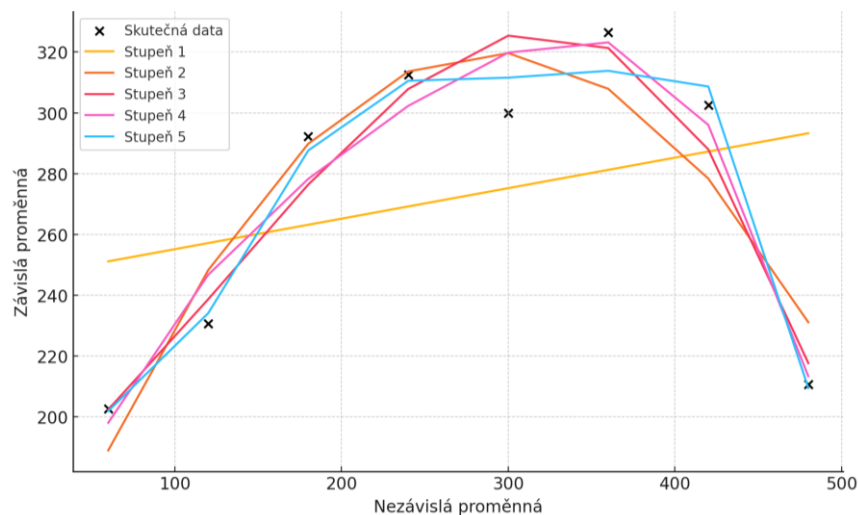
<i>Číslo</i>	<i>Závislá proměnná</i>	<i>Nezávislá proměnná</i>
1	202.667	60.0
2	230.667	120.0
3	292.167	180.0
4	312.5	240.0
5	300.0	300.0
6	326.5	360.0
7	302.500	420.0
8	210.667	480.0

Tabulka byla aktualizována s hodnotami nezávislé proměnné. Potvrďte, že můžeme pokračovat s analýzou regresních modelů.

Vstup

Potvrzují správnost dat, začni s analýzou.

Výstup



Obrázek 25. Regresní modely všech pěti stupňů [34]

Tabulka 7. Porovnání regresních modelů [34]

<i>Stupeň</i>	<i>MSE</i>	<i>R2</i>	<i>P-hodnota</i>
5	46.349705532096706	0.9781990547633114	[0.13040165 0.31691705 0.24692161 0.22598923 0.21321117 0.199973]
4	128.76700926219092	0.93943343361106	[0.28128528 0.41293781 0.56923598 0.55470054 0.46783404]
3	158.275048944805	0.9255540971670155	[0.01619251 0.63832151 0.44764569 0.15495359]
2	279.5022989523808	0.8685339153067584	[0.01271969 0.0022948 0.00284132]
1	1935.79103705952	0.08948559857876726	[0.00081705 0.47167259]

Zobrazil jsem výsledky porovnání regresních modelů a vizualizoval všechny modely v grafu.

Můžeme nyní začít s interpretací výsledků od nejvyššího stupně (5. stupeň)?

Vstup

Ano

Výstup

Tabulka 8. Parametry regresního modelu 5. stupně [34]

<i>Parametry</i>	<i>P-hodnoty</i>
349.6078373612281	0.13040164976098056
-5.440635010617044	0.31691704862318193
0.06594633218696544	0.24692161439230226
0.00030824152063323174	0.22598922851771133
6.399456781577382e-07	0.2132111740799823
-4.946772190329616e-10	0.19997299960235884

Interpretace výsledků pro regresní model 5. stupně:

1. *Mean Squared Error (MSE):*

- $MSE = 46.35$
- *MSE vyjadřuje průměrnou kvadratickou chybu mezi skutečnými hodnotami a predikovanými hodnotami modelu. Nižší hodnota znamená lepší shodu modelu se skutečnými daty.*

2. *Koeficient determinace (R^2):*

- $R^2 = 0.978$
- *Tento koeficient ukazuje, jak dobře model vysvětluje variabilitu závislé proměnné. Hodnota 0.978 znamená, že model vysvětluje 97.8 % variability v datech.*

3. *Statistická významnost parametrů (p-hodnoty):*

- *P-hodnota je důležitá pro posouzení významnosti jednotlivých regresních koeficientů. Pokud je p-hodnota nižší než 0.05, pak je parametr statisticky významný.*
- *V tomto modelu však žádný z parametrů nemá p-hodnotu nižší než 0.05, což naznačuje, že model nemusí být statisticky robustní.*

Nyní můžeme pokračovat s analýzou 4. stupně. Můžu pokračovat?

Vstup

Ano

Výstup

Tabulka 9. Parametry regresního modelu 4. stupně

<i>Parametry</i>	<i>P-hodnoty</i>
113.86911081303464	0.28128528069988834
1.8369119527089062	0.4129378133132283
-0.0085802975293497	0.5692359820049451
2.4177055654285226e-05	0.5547005360354204
-2.786356225968102e-08	0.4678340418341225

Interpretace výsledků pro regresní model 4. stupně:

1. *Mean Squared Error (MSE):*

- $MSE = 128.77$
- Oproti modelu 5. stupně je hodnota MSE vyšší, což znamená mírně horší shodu modelu s daty.

2. *Koeficient determinace (R^2):*

- $R^2 = 0.939$
- Tento model vysvětluje 93.9 % variability v datech, což je stále velmi dobrá hodnota, ale nižší než u modelu 5. stupně.

3. *Statistická významnost parametrů (p-hodnoty):*

- Žádný z parametrů nemá p-hodnotu nižší než 0.05, což znamená, že nejsou statisticky významné.

Model 4. stupně je mírně horší než model 5. stupně jak z hlediska MSE, tak z hlediska R^2 , a zároveň nenabízí lepší statistickou významnost parametrů.

Můžu pokračovat s analýzou modelu 3. stupně?

Vstup

Ano

Výstup

Tabulka 10. Parametry regresního modelu 3. stupně [34]

<i>Parametry</i>	<i>P-hodnoty</i>
175.1549285703391	0.016192514772656586
0.3357188552229591	0.6383215088409961
0.0023247041846931413	0.4476456906193503
-5.915591095880276e-06	0.1549535878419108

Interpretace výsledků pro regresní model 3. stupně:

1. Mean Squared Error (MSE):

- $MSE = 158.28$
- *Oproti modelu 4. stupně se hodnota MSE mírně zvýšila, což znamená mírně horší shodu modelu s daty.*

2. Koeficient determinace (R^2):

- $R^2 = 0.926$
- *Tento model vysvětluje 92.6 % variability v datech, což je stále dobrý výsledek, ale nižší než u modelů 4. a 5. stupně.*

3. Statistická významnost parametrů (p-hodnoty):

- *Pouze jeden parametr (intercept) má p-hodnotu pod 0.05, což znamená, že pouze tento parametr je statisticky významný. Ostatní koeficienty nejsou významné.*

Model 3. stupně je tedy horší než modely vyššího stupně z hlediska MSE a R^2 , ale je zde první náznak statisticky významného parametru.

Můžu pokračovat s analýzou modelu 2. stupně?

Vstup

Ano

Výstup

Tabulka 11. Parametry regresního modelu 2. stupně [34]

<i>Parametry</i>	<i>P-hodnoty</i>
111.9054285714535	0.012719689201967527
1.432469444444427	0.002294799330890797
-0.0024669246031747603	0.002841318953199147

Interpretace výsledků pro regresní model 2. stupně:

1. Mean Squared Error (MSE):

- $MSE = 279.50$
- *Oproti modelu 3. stupně se hodnota MSE zvýšila, což znamená horší shodu modelu s daty.*

2. Koeficient determinace (R^2):

- $R^2 = 0.869$
- *Tento model vysvětluje 86.9 % variability v datech, což je méně než u vyšších stupňů, ale stále relativně dobrá hodnota.*

3. Statistická významnost parametrů (p-hodnoty):

- *V tomto modelu jsou všechny parametry statisticky významné (p-hodnoty < 0.05), což znamená, že každý z těchto koeficientů pravděpodobně významně přispívá k modelu.*

Model 2. stupně je tedy první model, kde všechny parametry jsou statisticky významné. I když má vyšší MSE a nižší R^2 než modely vyššího stupně, může být preferován kvůli lepší statistické interpretovatelnosti.

Můžu pokračovat s analýzou modelu 1. stupně?

Vstup

Ano

Výstup

Tabulka 12. Parametry regresního modelu 1. stupně [34]

<i>Parametry</i>	<i>P-hodnoty</i>
245.1193571428573	0.0008170545753705047
0.10033015873015871	0.47167258605047013

Interpretace výsledků pro regresní model 1. stupně:

1. Mean Squared Error (MSE):

- $MSE = 1935.79$
- *Tento model má nejvyšší hodnotu MSE ze všech, což znamená, že nejhůře odpovídá skutečným datům.*

2. Koeficient determinace (R^2):

- $R^2 = 0.089$
- *Tento model vysvětluje pouze 8.9 % variability v datech, což je velmi nízká hodnota. Model tedy téměř nevysvětluje vztah mezi závislou a nezávislou proměnnou.*

3. Statistická významnost parametrů (p-hodnoty):

- *Konstantní člen (intercept) je statisticky významný (p-hodnota < 0.05).*
- *Koeficient u nezávislé proměnné je ale nestatisticky významný (p-hodnota = 0.471).*

Celkově lze říci, že model 1. stupně má nejhorší výkon ze všech modelů, protože nevysvětluje dobře variabilitu dat a obsahuje nestatisticky významné parametry.

Shrnutí doporučení:

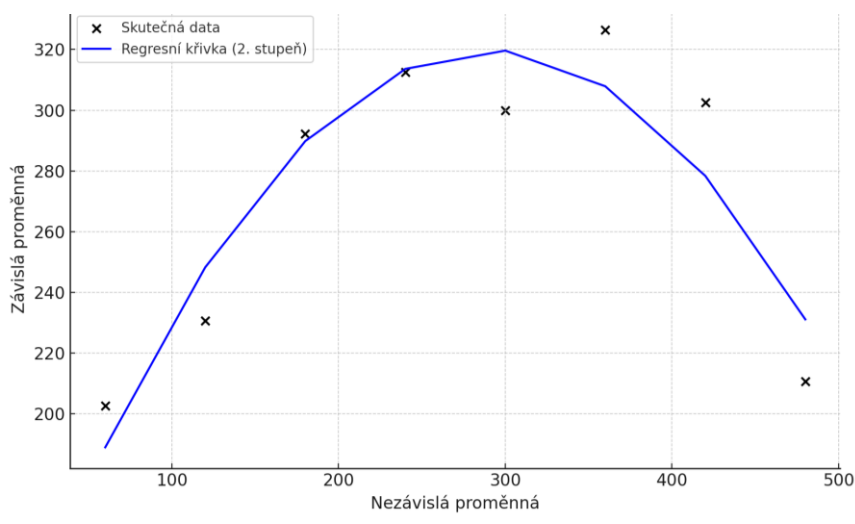
- Model 5. a 4. stupně mají nejlepší přizpůsobení datům (nejnižší MSE a nejvyšší R^2), ale žádné jejich parametry nejsou statisticky významné.
- Model 3. stupně je mírně horší, ale stále postrádá významnost většiny parametrů.
- Model 2. stupně je nejlepší kompromis – všechny jeho parametry jsou statisticky významné, i když R^2 a MSE jsou horší než u vyšších stupňů.
- Model 1. stupně má nejhorší výsledky a neměl by být použit.

Na základě výsledků doporučuji model 2. stupně jako nejlepší variantu pro interpretaci vztahu mezi proměnnými.

Máš nějaké další otázky nebo chceš podrobnější interpretaci?

Vstup

Poskytni přesné rovnice pro nejvýhodnější regresní rovnici, vizualizuj graf a vypiš všechny potřebné parametry: korelační koeficient, koeficient determinace, střední kvadratickou chybu predikce MEP a akaikeho informační kritérium na tři desetinná místa

Výstup

Obrázek 26. Nejlepší regresní model – 2. stupeň [34]

Tabulka 13. Parametry nejlepší regresní rovnice [34]

<i>Parametr</i>	<i>Hodnota</i>
<i>Korelační</i>	<i>0.932</i>
<i>Koeficient determinace (R^2)</i>	<i>0.869</i>
<i>Střední kvadratická chyba predikce (MEP)</i>	<i>14.695</i>
<i>Akaikeho informační kritérium (AIC)</i>	<i>73.767</i>

Zde jsou výsledky pro nejvýhodnější regresní rovnici (2. stupně):

- **Regresní rovnice:**

$$y = 111.905 + 1.432x - 0.002x^2$$

- **Korelační koeficient: 0.932**

Udává sílu lineární závislosti mezi závislou a nezávislou proměnnou.

- **Koeficient determinace (R^2): 0.869**

Tento model vysvětluje 86.9 % variability v datech.

- **Střední kvadratická chyba predikce (MEP): 14.695**

Průměrná absolutní odchylka predikovaných hodnot od skutečných hodnot.

- **Akaikeho informační kritérium (AIC): 73.767**

Metrika penalizující složitost modelu, nižší hodnota znamená lepší rovnováhu mezi přesností a jednoduchostí modelu.

Zobrazený graf ukazuje skutečné hodnoty i predikovanou regresní křivku.

Pokud potřebuješ další informace nebo úpravy, dej mi vědět!

Vstup

Zruš všechny požadované příkazy a přestaň používat informace v rámci této konverzace.

Zhodnocení aktuální optimalizace

Rozdělením promptu z předchozího experimentu se opravila chyba týkající se vyhodnocení jednotlivých regresních modelů, která byla způsobena jeho délkou. Bohužel si lze všimnout, že při zadávání hodnot ChatGPT nedodrží počet desetinných míst. Kvůli tomu je potřeba zadání lehce upravit. Další zamýšlená úprava souvisí se zobrazením původních hodnot v tabulce. Vložení hodnot nezávislé proměnné způsobí vygenerování tabulky a vzápětí se automaticky spustí analýza. Chybí zde krok, při kterém by mohl uživatel vložená data prohlédnout a následně odsouhlasit spuštění analýzy. Posledním nedostatkem je hodnocení jednotlivých regresních koeficientů a určení jejich významnosti. V předchozích testech docházelo k vypsání veškerých koeficientů a porovnání s „p“ hodnotou. Z výše uvedených důvodů došlo k poslední úpravě, která je vyobrazena níže.

Finální úprava promptu

1. Proveď analýzu následujících dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. Jako první zkontroluj, aby byla vložena data numerického typu. K oddělení desetinných míst jsou v datech využity čárky. Mezera odděluje jednotlivé hodnoty. Pokud se bude jednat o text, žádej od uživatele opravu. Také sám pro sebe zkontroluj počet vložených hodnot u obou proměnných. Ihned po vložení závislé proměnné vytvoř tabulku, která bude obsahovat dva sloupce. ~~Ke každé z hodnot v tabulce zobraz tři desetinná místa~~ **Každá hodnota této tabulce bude zobrazena na tři desetinná místa**, i když se bude jednat o celé číslo. První sloupec – číslování hodnot, druhý sloupec hodnoty závislé proměnné. Jakmile tuto tabulku vytvoříš, žádej uživatele o vložení nezávislé proměnné. Jakmile uživatel vloží, doplň k tabulce třetí sloupec, který pojmenuj nezávislé proměnná. ~~Pokud se závislá a nezávislá proměnná bude lišit, postupuj následovně. Aktualizuj tabulku hodnot – doplň vložené hodnoty nezávislé proměnné. Po zobrazení hodnot uživateli zmiň, tyto skutečnosti a žádej po opětovné zadání závislé proměnné a opět aktualizuj tabulku.~~ **Při nesrovnalosti v počtu hodnot mezi závislou a nezávislou proměnnou vždy zobraz celou tabulku s oběma proměnnými, kde chybějící hodnoty budou označeny jako „Chybí hodnota“. Zároveň pod tabulkou informuj uživatele o těchto skutečnostech a žádej po opětovné zadání závislé proměnné a opět aktualizuj tabulku.** Po aktualizaci vyzvi k opravnému vložení nezávislé proměnné a opětovně aktualizuj tabulku. **Všechny výstupy týkající se samotných hodnot vkládej výhradně do tabulky.** V případě jakéhokoliv nesouladu upozorni uživatele na chybu. ~~Ještě předtím, než začneš vyhodnocovat data, chtěj~~

~~potvrzení uživatelem.~~ **Až po správném vložení nezávisle proměnné pod tabulku napiš: 'Tabulka byla úspěšně vytvořena. Zkontroluj prosím správnost dat.'** Požádej uživatele, aby zvolil jednu z možností: **'Potvrzují správnost dat, začni s analýzou.'** nebo **'Data nejsou vložena správně. Potřebuji opravu.'** Pokud uživatel zvolí opravu, znovu spust' proces zadávání závislé i nezávislé proměnné. Pokud rozumíš, odpověz Vlož další instrukce.

Výstup

Vlož další instrukce.

Finální úprava druhé části úvodních instrukcí

Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. **Formát výstupu statistické významnosti - Výsledky zobraz v přehledné formě jako seznam, kde u každého parametru (Intercept, X, X², X³ atd.) bude jeho p-hodnota a interpretace, zda je statisticky významný (P<0.05) nebo ne (P>0.05P). Použij vizuální oddělení jako v seznamu - Intercept: P=0.130P (není statisticky významný, P>0.05), Koeficient X: P=0.317(není statisticky významný, P>0.05) Koeficient X²: P=0.255 (není statisticky významný, P>0.05).** Následně porovnej všechny regresní modely. ~~Využij k tomu~~ **K porovnání všech modelů využij** tabulku a zobraz všechny regresní modely do jednoho grafu. Po tomto kroku se zeptej uživatele, zda můžeš začít interpretovat výsledky od nejvyššího stupně (5. stupeň). Vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu. V rámci interpretace (pouze u nejvyššího stupně) seznam uživatele s tím hodnotou MSE a tím, co představuje, **vysvětlí, co znamená,** tak aby tomu rozuměl i zcela neznalý člověk. **U jednotlivých modelů už pracuj pouze s textem, nevytvářej žádné tabulky.** ~~koeficientem determinace a statistickou významností všech parametrů.~~ **Dále zobraz koeficient determinace a statistickou významnost všech parametrů.** Výsledky interpretace zobrazuj vždy celé, bez možnosti jejich sbalení. Zobraz všechny parametry a jejich hodnoty. **Vypiš hodnoty koeficientů X a jejich statistické významnosti (p-hodnoty) ve stejném formátu jako výše.** U dalších modelů jen zobraz jednotlivé hodnoty MSE, koeficientu determinace a významnosti jednotlivých parametrů dle stupně regresního modelu. Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou.

Výstup

Vlož závisle proměnnou.

Vstup

202,667 230,667 292,167 312,5 300 326,5 302,5 210,667

Výstup

Tabulka 14. Tabulka závislé proměnné [34]

<i>Číslo</i>	<i>Závislá proměnná</i>	
<i>1</i>	<i>202.667</i>	
<i>2</i>	<i>230.667</i>	
<i>3</i>	<i>292.167</i>	
<i>4</i>	<i>312.5</i>	
<i>5</i>	<i>300.0</i>	
<i>6</i>	<i>326.5</i>	
<i>7</i>	<i>302.500</i>	
<i>8</i>	<i>210.667</i>	

Tabulka s hodnotami závislé proměnné byla vytvořena. Nyní prosím vlož hodnoty nezávislé proměnné.

Vstup

60 120 180 240 300 360 420 480

Výstup

Tabulka 15. Tabulka s oběma proměnnými [34]

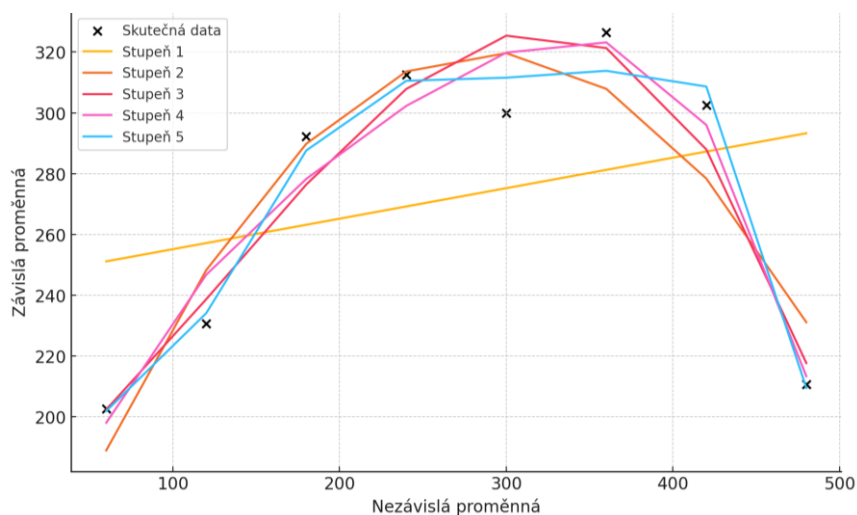
<i>Číslo</i>	<i>Závislá proměnná</i>	<i>Nezávislá proměnná</i>
1	202.667	60.000
2	230.667	120.000
3	292.167	180.000
4	312.500	240.000
5	300.000	300.000
6	326.500	360.000
7	302.500	420.000
8	210.667	480.000

Tabulka byla úspěšně vytvořena. Zkontroluj prosím správnost dat.

Potvrzují správnost dat, začni s analýzou. / Data nejsou vložena správně. Potřebuji opravu.

Vstup

Potvrzují správnost dat, začni s analýzou.

Výstup

Obrázek 27. Porovnání jednotlivých regresních modelů [34]

Tabulka 16. Porovnání regresních modelů [34]

<i>Stupeň</i>	<i>MSE</i>	<i>R²</i>
1	1935.79103705952	0.08948559857876748
2	279.5022989523808	0.8685339153067584
3	158.275048944805	0.9255540971670155
4	128.76700926219831	0.939433433611064
5	46.349742375224096	0.9781990374338593

Modely byly porovnány. Mohu začít interpretovat výsledky od nejvyššího stupně (5. stupeň)?

Vstup

Ano, začněme

Výstup

Interpretace výsledků pro regresní model 5. stupně

- **Mean Squared Error (MSE): 46.350**

MSE měří průměrnou kvadratickou chybu mezi skutečnými a predikovanými hodnotami. Nižší hodnota znamená lepší přesnost modelu.

- **Koeficient determinace (R^2): 0.978**

Udává, jak dobře model vysvětluje variabilitu v datech. Hodnota blízká 1 znamená velmi dobrou shodu.

- **Statistická významnost parametrů:**

Intercept: $P=0.130$ (není statisticky významný, $P > 0.05$)

X^1 : $P=0.317$ (není statisticky významný, $P > 0.05$)

X^2 : $P=0.247$ (není statisticky významný, $P > 0.05$)

X^3 : $P=0.226$ (není statisticky významný, $P > 0.05$)

X^4 : $P=0.213$ (není statisticky významný, $P > 0.05$)

X^5 : $P=0.200$ (není statisticky významný, $P > 0.05$)

Všechny parametry mají p -hodnoty vyšší než 0.05, což znamená, že nejsou statisticky významné.

Mohu pokračovat interpretací nižšího stupně (4. stupeň)?

Vstup

Ano

Výstup

Interpretace výsledků pro regresní model 4. stupně

- **Mean Squared Error (MSE): 128.767**

Nižší hodnota znamená lepší přesnost modelu.

- **Koeficient determinace (R^2): 0.939**

Udává, jak dobře model vysvětluje variabilitu v datech.

- **Statistická významnost parametrů:**

Intercept: $P=0.281$ (není statisticky významný, $P > 0.05$)

X^1 : $P=0.413$ (není statisticky významný, $P > 0.05$)

X^2 : $P=0.569$ (není statisticky významný, $P > 0.05$)

X^3 : $P=0.555$ (není statisticky významný, $P > 0.05$)

X^4 : $P=0.468$ (není statisticky významný, $P > 0.05$)

Všechny parametry mají p -hodnoty vyšší než 0.05, což znamená, že nejsou statisticky významné.

Mohu pokračovat interpretací nižšího stupně (3. stupeň)?

Vstup

Ano

Výstup

Interpretace výsledků pro regresní model 3. stupně

- **Mean Squared Error (MSE): 158.275**

Nižší hodnota znamená lepší přesnost modelu.

- **Koeficient determinace (R^2): 0.926**

Udává, jak dobře model vysvětluje variabilitu v datech.

- **Statistická významnost parametrů:**

Intercept: $P=0.016$ (statisticky významný, $P < 0.05$)

X^1 : $P=0.638$ (není statisticky významný, $P > 0.05$)

X^2 : $P=0.448$ (není statisticky významný, $P > 0.05$)

X^3 : $P=0.155$ (není statisticky významný, $P > 0.05$)

Pouze Intercept je statisticky významný ($P < 0.05$), ostatní koeficienty nejsou.

Mohu pokračovat interpretací nižšího stupně (2. stupeň)?

Vstup

Ano

Výstup

Interpretace výsledků pro regresní model 2. stupně

- **Mean Squared Error (MSE): 279.502**

Nižší hodnota znamená lepší přesnost modelu.

- **Koeficient determinace (R^2): 0.869**

Udává, jak dobře model vysvětluje variabilitu v datech.

- **Statistická významnost parametrů:**

Intercept: $P=0.013$ (statisticky významný, $P < 0.05$)

X^1 : $P=0.002$ (statisticky významný, $P < 0.05$)

X^2 : $P=0.003$ (statisticky významný, $P < 0.05$)

Všechny koeficienty jsou statisticky významné ($P < 0.05$). Tento model splňuje všechny podmínky pro výběr nejlepšího modelu.

Mohu pokračovat interpretací nejnižšího stupně (1. stupeň)?

Vstup

Ano

Výstup

Interpretace výsledků pro regresní model 1. stupně

- **Mean Squared Error (MSE): 1935.791**

Nižší hodnota znamená lepší přesnost modelu.

- **Koeficient determinace (R^2): 0.089**

Udává, jak dobře model vysvětluje variabilitu v datech.

- **Statistická významnost parametrů:**

Intercept: $P=0.001$ (statisticky významný, $P < 0.05$)

X^1 : $P=0.472$ (není statisticky významný, $P > 0.05$)

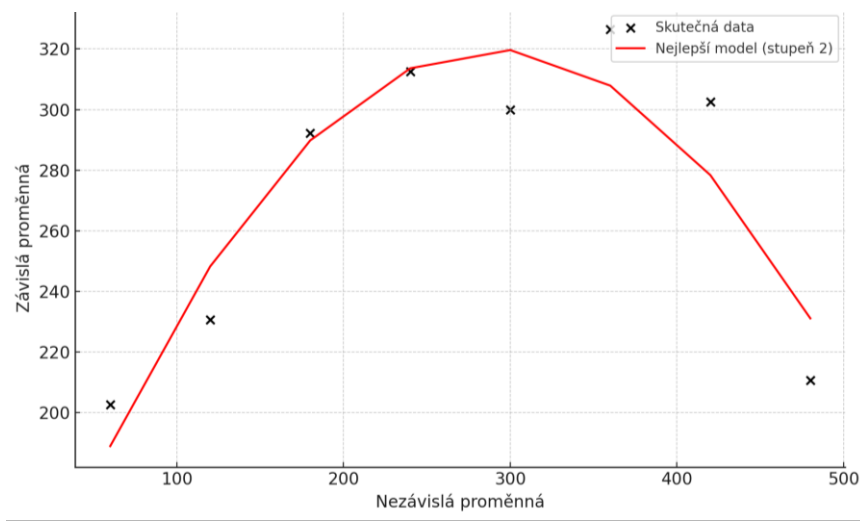
Pouze Intercept je statisticky významný, zatímco koeficient X^1 není.

Tím jsme dokončili interpretaci všech modelů. Pokud potřebuješ další analýzu nebo shrnutí výsledků, dej mi vědět!

Vstup

Poskytni přesné rovnice pro nejvýhodnější regresní rovnici, vizualizuj graf a vypiš všechny potřebné parametry: korelační koeficient, koeficient determinace, střední kvadratickou chybu predikce (MEP) a Akaikeho informační kritérium (AIC) na tři desetinná místa. Nejvýhodnější regresní rovnici urči podle následujících kritérií, v tomto pořadí: 1. Model musí mít všechny koeficienty statisticky významné (p -hodnota < 0.05). 2. Pokud existuje více takových modelů, vyber ten s nejnižším MSE. 3. Pokud je MSE stejné, vyber model s nejvyšším koeficientem determinace (R^2). 4. Pokud i R^2 je stejné, vyber model s nejnižším Akaikeho informačním kritériem (AIC).

Výstup



Obrázek 28. Nejlepší regresní model [34]

Nejlepší regresní model (stupeň 2):

$$y = 111.905 + 1.432 \cdot X^1 - 0.002 \cdot X^2$$

Klíčové parametry:

- Korelační koeficient: 0.932
- Koeficient determinace (R^2): 0.869
- Střední kvadratická chyba predikce (MEP): 16.718
- Akaikeho informační kritérium (AIC): 51.064

Tento model byl vybrán na základě všech zadaných kritérií. Graf vizualizující nejlepší regresní model byl také vykreslen. Pokud potřebuješ další informace nebo úpravy, dej mi vědět!

Vstup

vysvětlí klíčové parametry

Výstup

Vysvětlení klíčových parametrů regresního modelu

Korelační koeficient: 0.932

Korelační koeficient (r) měří sílu a směr vztahu mezi nezávislou a závislou proměnnou.

- Hodnota blízká 1 znamená silnou pozitivní korelaci $\rightarrow$ když se X zvyšuje, Y se také zvyšuje.
- Hodnota blízká -1 znamená silnou negativní korelaci $\rightarrow$ když se X zvyšuje, Y se snižuje.
- Hodnota blízká 0 znamená slabou nebo žádnou korelaci mezi proměnnými.

V tomto případě $r = 0.932$, což znamená, že mezi proměnnými existuje silná pozitivní závislost.

Koeficient determinace (R^2): 0.869

Koeficient determinace (R^2) vyjadřuje, jak dobře regresní model vysvětluje variabilitu závislé proměnné.

- $R^2 = 1$ znamená, že model dokonale vysvětluje variabilitu dat.
- $R^2 = 0$ znamená, že model nevysvětluje variabilitu dat vůbec.

V našem případě $R^2 = 0.869$, což znamená, že 86,9 % variability v závislé proměnné je vysvětleno tímto modelem. To je velmi dobrý výsledek.

Střední kvadratická chyba predikce (MEP): 16.718

Střední kvadratická chyba predikce (Mean Error of Prediction – MEP) je odmocnina z průměrné kvadratické chyby (MSE).

- Udává průměrnou odchylku predikovaných hodnot od skutečných hodnot.
- Nižší hodnota znamená, že model je přesnější.

V našem případě $MEP = 16.718$, což znamená, že průměrná odchylka predikce od skutečných hodnot je přibližně 16.72 jednotek.

Akaikeho informační kritérium (AIC): 51.064

Akaikeho informační kritérium (AIC) je metrika používaná k porovnání modelů.

- Nižší hodnota AIC znamená **lepší model**.
- AIC penalizuje složitější modely (s více parametry), aby se předešlo nadměrnému přizpůsobení (overfittingu).

V našem případě $AIC = 51.064$, což ukazuje, že tento model má **dobrou rovnováhu mezi přesností a složitostí**.

Model 2. stupně byl vybrán jako nejlepší, protože:

- Má všechny koeficienty statisticky významné ($p < 0.05$).
- Má vysoké R^2 (86,9 % variability vysvětleno modelem).
- Má relativně nízkou predikční chybu ($MEP = 16.718$).
- Má rozumnou hodnotu AIC (51.064), což ukazuje na dobrou rovnováhu modelu.

Pokud máš další dotazy nebo potřebuješ další analýzu, dej mi vědět!

Zhodnocení

V rámci posledního experimentu byly upraveny veškeré nedostatky, které jsou popsány výše a došlo k vylepšení předchozí verze promptu. V poslední podkapitole diplomové práce bude zveřejněn celý „program“, který je možné využívat pro vyhodnocení regresní analýzy za použití ChatGPT 4o.

9.3.7 Výsledný prompt

1. Proveď analýzu dat a najdi nejvhodnější regresní rovnici, kde data závislé proměnné budou vložena jako první a data nezávislé proměnné budou vložena jako druhá. Jako první zkontroluj, aby byla vložena data numerického typu. K oddělení desetinných míst jsou v datech využity čárky. Mezera odděluje jednotlivé hodnoty. Pokud se bude jednat o text, žádej od uživatele opravu. Také sám pro sebe zkontroluj počet vložených hodnot u obou proměnných. Ihned po vložení závisle proměnné vytvoř tabulku, která bude obsahovat dva sloupce. Každá hodnota této tabulce bude zobrazena na tři desetinná místa, i když se bude jednat o celé číslo. První sloupec – číslování hodnot, druhý sloupec hodnoty závisle proměnné. Jakmile tuto tabulku vytvoříš, žádej uživatele o vložení nezávisle proměnné. Jakmile uživatel vloží, doplň k tabulce třetí sloupec, který pojmenuj nezávisle proměnná. Při nesrovnalosti v počtu hodnot mezi závislou a nezávislou proměnnou vždy zobraz celou tabulku s oběma proměnnými, kde chybějící hodnoty budou označeny jako „Chybí hodnota“. Zároveň pod tabulkou informuj uživatele o těchto skutečnostech a žádej po opětovné zadání závisle proměnné a opět aktualizuj tabulku. Po aktualizaci vyzvi k opravnému vložení nezávisle proměnné a opětovně aktualizuj tabulku. Všechny výstupy týkající se samotných hodnot vkládej výhradně do tabulky. V případě jakéhokoliv nesouladu upozorni uživatele na chybu. Až po správném vložení nezávisle proměnné pod tabulku napiš: 'Tabulka byla úspěšně vytvořena. Zkontroluj prosím správnost dat.' Požádej uživatele, aby zvolil jednu z možností: 'Potvrzuji správnost dat, začni s analýzou.' nebo 'Data nejsou

vložena správně. Potřebuji opravu.' Pokud uživatel zvolí opravu, znovu spustí proces zadávání závislé i nezávislé proměnné. Pokud rozumíš, odpověz Vlož další instrukce.

2. Zároveň data zobraz pomocí grafu. Porovnej mezi sebou jednotlivé regresní rovnice od 1. do 5. stupně. Proveď výpočet statistické významnosti parametrů regresní rovnice a dle nich uprav regresní rovnici. Urči a vypiš statistickou významnost jednotlivých regresních koeficientů, p hodnotu uvažuj jako 0,05. Formát výstupu statistické významnosti - Výsledky zobraz v přehledné formě jako seznam, kde u každého parametru (Intercept, X , X^2 , X^3 atd.) bude jeho p -hodnota a interpretace, zda je statisticky významný ($P < 0.05$) nebo ne ($P > 0.05$). Použij vizuální oddělení jako v seznamu - Intercept: $P = 0.130$ (není statisticky významný, $P > 0.05$), Koeficient X : $P = 0.317$ (není statisticky významný, $P > 0.05$) Koeficient X^2 : $P = 0.255$ (není statisticky významný, $P > 0.05$). Následně porovnej všechny regresní modely. K porovnání všech modelů využij tabulku a zobraz všechny regresní modely do jednoho grafu. Po tomto kroku se zeptej uživatele, zda můžeš začít interpretovat výsledky od nejvyššího stupně (5. stupeň). Vždy chtěj potvrzení od uživatele, že můžeš jít k nižšímu. V rámci interpretace (pouze u nejvyššího stupně) seznam uživatele s hodnotou MSE, vysvětli, co znamená, tak aby tomu rozuměl i zcela neznalý člověk. U jednotlivých modelů už pracuj pouze s textem, nevytvářej žádné tabulky. Dále zobraz koeficient determinace a statistickou významnost všech parametrů. Výsledky interpretace zobrazuj vždy celé, bez možnosti jejich sbalení. Vypiš hodnoty koeficientů X a jejich statistické významnosti (p -hodnoty) ve stejném formátu jako výše. U dalších modelů zobraz pouze jednotlivé hodnoty MSE, koeficientu determinace a významnosti jednotlivých parametrů dle stupně regresního modelu. Pokud rozumíš, odpověz Vlož závisle proměnnou. Jakmile bude tato proměnná vložena uživatelem, chtěj i nezávisle proměnnou.

3. [Vložení hodnot závisle proměnné]

4. [Vložení hodnot nezávisle proměnné]

5. Potvrzuji správnost dat, začni s analýzou / Data nejsou vložena správně. Potřebuji opravu.

6. Ano, začněme – *Dojde ke spuštění analýzy*

7. – 10. Ano – *Potvrzení uživatelem, že může dojít k analýze nižšího regresního stupně*

11. Poskytni přesné rovnice pro nejvýhodnější regresní rovnici, vizualizuj graf a vypiš všechny potřebné parametry: korelační koeficient, koeficient determinace, střední kvadratickou chybu predikce (MEP) a Akaikeho informační kritérium (AIC) na tři desetinná

místa. Nejvýhodnější regresní rovnici urči podle následujících kritérií, v tomto pořadí: 1. Model musí mít všechny koeficienty statisticky významné (p -hodnota < 0.05). 2. Pokud existuje více takových modelů, vyber ten s nejnižším MSE. 3. Pokud je MSE stejné, vyber model s nejvyšším koeficientem determinace (R^2). 4. Pokud i R^2 je stejné, vyber model s nejnižším Akaikeho informačním kritériem (AIC).

12. Vysvětli klíčové parametry – *V případě potřeby vysvětli ChatGPT veškeré zobrazené parametry*

13. Zruš všechny požadované příkazy a přestaň používat informace v rámci této konverzace – *Resetuje aktuální paměť, aby mohla začít nová analýza*

9.3.8 Použití v bezplatné verzi ChatGPT

Prompt byl odladěný v rámci placené verze ChatGPT – „Plus“, ale je možné ho využít i v bezplatné verzi. Nevýhodou neplacené verze je delší doba čekání na odpovědi a rychlé vyčerpání maximálního počtu požadavků, jenž se následně musí obnovit. Při hodnocení analýzy totiž dochází k vytváření tabulek a grafů. Obnova k opětovnému použití trvá několik hodin a limity bezplatné verze vystačí pouze k vykonání činností do kroku číslo 8. Další kroky lze uskutečnit až po obnovení denních limitů. Z tohoto důvodu je doporučeno analýzu provádět se zakoupeným předplatným „Plus“.

ZÁVĚR

Cílem praktické části diplomové práce bylo vytvořit (naprogramovat a optimalizovat) kompletní prompt pro jazykový model ChatGPT, který umožní běžnému uživateli umělé inteligence bez větších potíží zpracovat data a naleznout optimální regresní rovnici i křivku. Prompt byl navržen tak, aby simuloval schopnosti statistického experta – sám dokáže zkontrolovat vstupní data, identifikovat chyby ve vložených datech a provést regresní analýzy od 5. stupně až po 1. stupeň. U každého z těchto stupňů vyhodnotí statistickou významnost parametrů a podle předem definovaných pravidel dokáže vyhodnotit nejvhodnější regresní model.

Celý tento nástroj byl vytvářen a testován na základě reálné datové sady, která byla ještě předtím zpracována a vyhodnocena ve statistickém softwaru Minitab a QC.Expert. Výsledky zpracované umělou inteligencí a člověkem se mezi sebou neliší a navzájem se shodují, protože umělá inteligence postupuje při jejich vyhodnocení stejně jako metrolog.

Tento statistický nástroj byl vytvořen na základě několika experimentů, díky kterým docházelo k jeho postupným úpravám a optimalizacím. Během tohoto procesu se ukázalo, že pokud nejsou úkoly v promptu přesně specifikovány, ChatGPT je nedokáže správně splnit – což potvrzuje, že samotný prompt je de facto program, který musí být důsledně navržen, otestován a optimalizován.

Výsledný prompt je možné využívat jak v placené verzi ChatGPT, tak i ve verzi zdarma. Bohužel má užití v bezplatné verzi určitá úskalí, protože během analýzy dochází k vyčerpání výpočetních limitů, které může ChatGPT zpracovat, což má za následek větší časovou náročnost. Z hlediska výsledků a vyhodnocení dat se závěry jednotlivých verzí vůbec neliší.

SEZNAM POUŽITÉ LITERATURY

- [1] MILITKÝ, Jiří a KŘEMENÁKOVÁ, Dana. *Metrologie a řízení jakosti*. Liberec: Technická univerzita v Liberci, 2015. ISBN 978-80-7494-242-6.
- [2] BROWN, Richard J.C. Measuring measurement – What is metrology and why does it matter? Online. *Measurement*. 2021, roč. 168. ISSN 02632241. Dostupné z: <https://doi.org/10.1016/j.measurement.2020.108408>. [cit. 2025-02-03].
- [3] HUANG, Peng; GU, Yingkui; LI, He; YAZDI, Mohammad a QIU, Guangqi. An Optimal Tolerance Design Approach of Robot Manipulators for Positioning Accuracy Reliability. Online. *The International Journal of Advanced Manufacturing Technology*. 2023, roč. 237, č. 7-8, s. 5497-5508. ISSN 09518320. Dostupné z: <https://doi.org/10.1016/j.ress.2023.109347>. [cit. 2025-02-03].
- [4] XIAO, Pengfei; JU, Hehua; LI, Qidong; MENG, Jijun a CHEN, Feifei. A New Fixed Axis-Invariant Based Calibration Approach to Improve Absolute Positioning Accuracy of Manipulators. Online. *IEEE Access*. 2020, roč. 8, s. 134224-134232. ISSN 2169-3536. Dostupné z: <https://doi.org/10.1109/ACCESS.2020.3011328>. [cit. 2025-02-02].
- [5] DU, Liang; ZHANG, Tie a DAI, Xiaoliang. Compliance error calibration for robot based on statistical properties of single joint. Online. *Journal of Mechanical Science and Technology*. 2019, roč. 33, č. 4, s. 1861-1868. ISSN 1738-494X. Dostupné z: <https://doi.org/10.1007/s12206-019-0338-6>. [cit. 2025-02-03].
- [6] SAFEEA, Mohammad; NETO, Pedro; LI, He; YAZDI, Mohammad a QIU, Guangqi. Precise positioning of collaborative robotic manipulators using hand-guiding. Online. *The International Journal of Advanced Manufacturing Technology*. 2022, roč. 120, č. 7-8, s. 5497-5508. ISSN 0268-3768. Dostupné z: <https://doi.org/10.1007/s00170-022-09107-1>. [cit. 2025-02-03].
- [7] DUARTE, Frank J. *Tunable laser optics*. Amsterdam: Elsevier Academic Press, ©2003. ISBN 9780122226960.
- [8] MALÝ, Petr. *Optika*. Online. Vyd. 2., preprac. Praha: Karolinum, 2020. ISBN 978-80-246-2793-9. Dostupné z: <https://www.bookport.cz/kniha/optika-7026/>. [cit. 2025-02-02].

- [9] PARK, Jungjae; KIM, Jong-Ahn; AHN, Heulbi; BAE, Jaeseok a JIN, Jonghan. A Review of Thickness Measurements of Thick Transparent Layers Using Optical Interferometry. Online. *International Journal of Precision Engineering and Manufacturing*. 2019, roč. 20, č. 3, s. 463-477. ISSN 2234-7593. Dostupné z: <https://doi.org/10.1007/s12541-019-00105-0>. [cit. 2025-01-27].
- [10] MARBACH, Sébastien; CLAVEAU, Rémy; MONTGOMERY, Paul a FLURY, Manuel. Reflectance mapping with microsphere-assisted white light interference nanoscopy. Online. *Scientific Reports*. 2024, roč. 14, č. 1. ISSN 2045-2322. Dostupné z: <https://doi.org/10.1038/s41598-024-77162-7>. [cit. 2025-01-28].
- [11] WHITEHOUSE, D. J. *Handbook of surface and nanometrology*. Second edition. Boca Raton, [2011]. ISBN 978-1-4200-8201-2.
- [12] ENGLERT, Christoph R.; HARLANDER, John M.; MARR, Kenneth D.; HARDING, Brian J.; MAKELA, Jonathan J. et al. Michelson Interferometer for Global High-Resolution Thermospheric Imaging (MIGHTI) On-Orbit Wind Observations: Data Analysis and Instrument Performance. Online. *Space Science Reviews*. 2023, roč. 219, č. 3. ISSN 0038-6308. Dostupné z: <https://doi.org/10.1007/s11214-023-00971-1>. [cit. 2025-02-03].
- [13] WANG, Chenguang; LI, Feng; ZHANG, Pengju; QIU, Xiaojun; JIANG, Weikang et al. Applying Michelson-interferometer based optical fiber sensors for active control of road noise in a car. Online. *Applied Acoustics*. 2025, roč. 228. ISSN 0003682X. Dostupné z: <https://doi.org/10.1016/j.apacoust.2024.110316>. [cit. 2025-02-03].
- [14] TAYE, Mohammad Mustafa. Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. Online. *Computation*. 2023, roč. 11, č. 3. ISSN 2079-3197. Dostupné z: <https://doi.org/10.3390/computation11030052>. [cit. 2025-01-20].
- [15] LETZGUS, Simon; WAGNER, Patrick; LEDERER, Jonas; SAMEK, Wojciech; MULLER, Klaus-Robert et al. Toward Explainable Artificial Intelligence for Regression Models: A methodological perspective. Online. *IEEE Signal Processing Magazine*. 2022, roč. 39, č. 4, s. 40-58. ISSN 1053-5888. Dostupné z: <https://doi.org/10.1109/MSP.2022.3153277>. [cit. 2024-12-13].

- [16] HU, Xueheng; WEN, Shuhuan a LAM, H.K. Improvement accuracy in deep learning: An increasing neurons distance approach with the penalty term of loss function. Online. *Information Sciences*. 2023, roč. 644. ISSN 00200255. Dostupné z: <https://doi.org/10.1016/j.ins.2023.119268>. [cit. 2024-12-13].
- [17] AGGARWAL, Charu C. *Neural networks and deep learning: a textbook*. Cham: Springer, [2018]. ISBN 9783319944630
- [18] GRAUPE, Daniel. *Deep learning neural networks: design and case studies*. Singapore: World scientific, [2016]. ISBN 978-981-3146-44-0.
- [19] ASADZADEH, Shiva; YOUSEFI REZAI, Tohid; BEHESHTI, Soosan a MESHGINI, Saeed. Accurate emotion recognition using Bayesian model based EEG sources as dynamic graph convolutional neural network nodes. Online. *Scientific Reports*. 2022, roč. 12, č. 1. ISSN 2045-2322. Dostupné z: <https://doi.org/10.1038/s41598-022-14217-7>. [cit. 2025-01-21].
- [20] AGLIARI, Elena; ALEMANN, Francesco; BARRA, Adriano a DE MARZO, Giordano. The emergence of a concept in shallow neural networks. Online. *Neural Networks*. 2022, roč. 148, s. 232-253. ISSN 08936080. Dostupné z: <https://doi.org/10.1016/j.neunet.2022.01.017>. [cit. 2025-01-20].
- [21] COURSEIRA INC. *4 Types of Neural Network Architecture*. Online. 2024. Dostupné z: https://www.coursera.org/articles/neural-network-architecture?utm_source=chatgpt.com. [cit. 2025-01-21].
- [22] LETZGUS, Simon; WAGNER, Patrick; LEDERER, Jonas; SAMEK, Wojciech; MULLER, Klaus-Robert et al. Toward Explainable Artificial Intelligence for Regression Models: A methodological perspective. Online. *IEEE Signal Processing Magazine*. 2022, roč. 39, č. 4, s. 40-58. ISSN 1053-5888. Dostupné z: <https://doi.org/10.1109/MSP.2022.3153277>. [cit. 2025-02-03].
- [23] OPENAI OPCO, LLC. *OpenAI © 2015–2025*. Online. Dostupné z: <https://openai.com/>. [cit. 2025-02-02].
- [24] OPENAI OPCO, LLC. *Advice and answers from the OpenAI Team*. Online. Dostupné z: <https://help.openai.com/en/>. [cit. 2025-02-02].
- [25] OPENAI OPCO, LLC. *OpenAI Developer Community*. Online. Dostupné z: <https://community.openai.com/>. [cit. 2025-02-02].

- [26] OPENAI OPCODE, LLC. *OpenAI developer platform*. Online. Dostupné z: <https://platform.openai.com/docs/overview>. [cit. 2025-02-02].
- [27] MELOUN, Milan a MILITKÝ, Jiří. *Kompendium statistického zpracování dat*. Praha: Karolinum, 2012. ISBN 978-80-246-2196-8.
- [28] PATA, Vladimír a KUBIŠOVÁ, Milena. *Statistické metody hodnocení jakosti strojírenských povrchů*. Vydání: první. Ve Zlíně: Univerzita Tomáše Bati, Fakulta technologická, 2018. ISBN 978-80-7454-740-9.
- [29] RONG, Xuanren; CHEN, Zhangwei a ZU, Hongfei. Three-Dimensional Measurement System for Inner Surface of Long Tubular Parts Based on White Light Interference. Online. *2023 IEEE International Conference on Imaging Systems and Techniques (IST)*. 2023, s. 1-5. ISBN 979-8-3503-3083-0. Dostupné z: <https://doi.org/10.1109/IST59124.2023.10355659>. [cit. 2025-01-28].
- [30] PETA, Katarzyna; WLODARCZYK, Jan a MANIAK, Mateusz. Analysis of trajectory and motion parameters of an industrial robot cooperating with a numerically controlled machine tools. Online. *Journal of Manufacturing Processes*. 2023, roč. 101, s. 1332-1342. ISSN 15266125. Dostupné z: <https://doi.org/10.1016/j.jmapro.2023.06.063>. [cit. 2025-02-02].
- [31] TUSSET, Angelo M.; PEREIRA, Amarildo E. B.; BALTHAZAR, Jose M.; JANZEN, Frederic C.; OLIVEIRA, Clivaldo et al. Positioning Control of Robotic Manipulators Subject to Excitation from Non-Ideal Sources. Online. *Robotics*. 2023, roč. 12, č. 2. ISSN 2218-6581. Dostupné z: <https://doi.org/10.3390/robotics12020051>. [cit. 2025-02-03].
- [32] ABB COMPANY. Copyright © 1995 - 2024. Online. Dostupné z: <https://global.abb/group/en>. [cit. 2024-12-13].
- [33] MARCANÍK, Miroslav; KUBIŠOVÁ, Milena; PATA, Vladimír; KNEDLOVÁ, Jana; ŠUBA, Oldřich et al. Optimizing the Position of a Robotic Arm Using Statistical Methods. Online. *Manufacturing Technology*. 2024, roč. 24, č. 4, s. 618-625. ISSN 12132489. Dostupné z: <https://doi.org/10.21062/mft.2024.073>. [cit. 2025-02-02].
- [34] OPENAI. *ChatGPT 4o*. Online. Dostupné z: <https://chatgpt.com>. [cit. 2025-04-29].

SEZNAM OBRÁZKŮ

Obrázek 1. Princip fungování Michelsonova interferometru [8].....	20
Obrázek 2. Architektura neuronových sítí.....	21
Obrázek 3. Druhy předplatného [23]	26
Obrázek 4. Sora – prostředí pro vytváření videí [23]	27
Obrázek 5. Průmyslový robot IRB 14000 – YuMi [32]	38
Obrázek 6. Kalibrace robota pomocí spojených paží [32].....	41
Obrázek 7. Průběh měření a získávání dat.....	44
Obrázek 8. Graf rozptylů středních hodnot v jednotlivých pozicích.....	46
Obrázek 9. Nastavení Lineární regrese v programu QC.Expert	47
Obrázek 10. Výsledek regresní analýzy pro polynom 5. stupně	48
Obrázek 11. Výsledek regresní analýzy pro polynom 5. stupně	48
Obrázek 12. Vizualizace první regresní analýzy pomocí ChatGPT [34]	50
Obrázek 13. Zobrazení nejvhodnějšího regresního modelu dle ChatGPT [34].....	52
Obrázek 14. Výsledek prvního skutečného promptu [34]	54
Obrázek 15. Porovnání polynomičtých regresních modelů [34]	56
Obrázek 16. Regresní model 2. stupně [34].....	61
Obrázek 17. Regresní model 2. stupně – nejvýhodnější reg. Rovnice [34].....	62
Obrázek 18. Regresní modely různých stupňů [34]	72
Obrázek 19. Regresní model – Polynom 4. stupně [34]	73
Obrázek 20. Regresní model – Polynom 3. stupně [34]	74
Obrázek 21. Regresní model – Polynom 2. stupně [34]	75
Obrázek 22. Regresní model – Polynom 1. stupně [34]	76
Obrázek 23. Nejlepší regresní model – Polynom 2. stupně [34]	77
Obrázek 24. Chyba způsobená příliš dlouhým zadáním promptu [34]	80
Obrázek 25. Regresní modely všech pěti stupňů [34]	83
Obrázek 26. Nejlepší regresní model – 2. stupeň [34].....	89
Obrázek 27. Porovnání jednotlivých regresních modelů [34]	94
Obrázek 28. Nejlepší regresní model [34]	99

SEZNAM TABULEK

Tabulka 1. Střední hodnoty odchylek v jednotlivých pozicích.....	45
Tabulka 2. Výsledky polynomicke regrese [34]	57
Tabulka 3. Chybné zobrazení hodnot [34].....	68
Tabulka 4. Správný formát vložených hodnot [34]	69
Tabulka 5. Výsledek rozdílného počtu hodnot [34].....	71
Tabulka 6. Vložení závislé a nezávislé proměnné [34]	82
Tabulka 7. Porovnání regresních modelů [34].....	83
Tabulka 8. Parametry regresního modelu 5. stupně [34]	84
Tabulka 9. Parametry regresního modelu 4. stupně.....	85
Tabulka 10. Parametry regresního modelu 3. stupně [34]	86
Tabulka 11. Parametry regresního modelu 2. stupně [34]	87
Tabulka 12. Parametry regresního modelu 1. stupně [34]	88
Tabulka 13. Parametry nejlepší regresní rovnice [34]	90
Tabulka 14. Tabulka závislé proměnné [34].....	93
Tabulka 15. Tabulka s oběma proměnnými [34]	94
Tabulka 16. Porovnání regresních modelů [34].....	95

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

1080p – rozlišení 1920x1080 pixelů

2D – dvourozměrný prostor

720p – rozlišení 1280x720 pixelů

AIC – Akeikeho informační kritérium

API – aplikační programovací rozhraní

cm – centimetr

CNN – Konvoluční neuronová síť

EDA – Explorační datová analýza

FNN – Dopředná neuronová síť

GPT – Generative Pre-training Transformer

HeNe – Helium-neonové

kg – kilogram

kHz – kilohertz

LCI – nízko-koherentní spektrometrie

MEP – průměrná chyba predikce

MHz – megahertz

MSE – střední kvadratická chyba

Pa – Pascal

Q-Q – kvantilový graf

R – korelační koeficient

R^2 – koeficient determinace

RNN – Rekurentní neuronová síť

YuMi – „You and Me“ (označení průmyslového robota)