

# ANALÝZA PREDIKČNÍCH FAKTORŮ DIABETU 2. TYPU POMOCÍ LOGISTICKÉ REGRESE

## TEORETICKÁ ČÁST

Diabetes mellitus je chronická metabolická porucha a jedna z nejrychleji rostoucích zdravotních krizí současnosti, která představuje vážné a život ohrožující onemocnění. Rizikovost diabetu spočívá především v jeho dlouhé bezpříznakové fázi, kvůli níž zůstává zhruba 50 % nemocných zcela nediagnostikováno (WHO, 2023).

Klíčem k odvrácení následků tohoto onemocnění je včasná diagnóza, která se opírá o správné vyhodnocení varovných příznaků nemoci. Mezi nejběžnější symptomy patří polyurie (časté močení), polydipsie (nadměrná žízeň) a polyfagie (nadměrný hlad). Mezi další časté projevy se řadí náhlý úbytek hmotnosti, obezita, celková slabost, zpomalené hojení ran, rozmazané vidění, nepříjemné svědění, podrážděnost nebo svalová ztuhlost (American Diabetes Association, 2024).

## VÝZKUMNÁ ČÁST

Základní hypotézou pro tuto práci je, že přítomnost specifických symptomů, popř. demografických ukazatelů, dokáže předpovědět, zda je u jedince přítomný diabetes.

**H: Existuje statisticky významný vztah mezi souborem sledovaných symptomů (věk, pohlaví, polyurie, polydipsie atd.) a výslednou diagnózou diabetu.**

Pro tuto práci byl použit dataset o 520 participantech s diabetem i bez diabetu, u kterých se buď vyskytují nebo nevyskytují nejrůznější symptomy typické pro diabetes. V tomto výzkumném souboru je 328 participantů mužů a 172 žen (dostupné na: <https://archive.ics.uci.edu/dataset/529/early+stage+diabetes+risk+prediction+dataset>).

Popis proměnných datového souboru:

*Nezávislé proměnné:*

- Věk
- Pohlaví
- Polyurie (časté močení)
- Polydipsie (nadměrná žízeň)

- Náhlý úbytek váhy
- Slabost
- Polyfagie (nadměrný hlad)
- Kvasinková infekce pohlavních orgánů
- Rozmazané vidění
- Svědění
- Podrážděnost
- Zpomalené hojení ran
- Parciální paréza (částečné snížení svalové hybnosti)
- Ztuhlost svalů
- Alopecie
- Obezita

*Závislá proměnná:*

- Diabetes:  
Přítomný – 1  
Nepřítomný – 0

### Výsledky logistické regrese

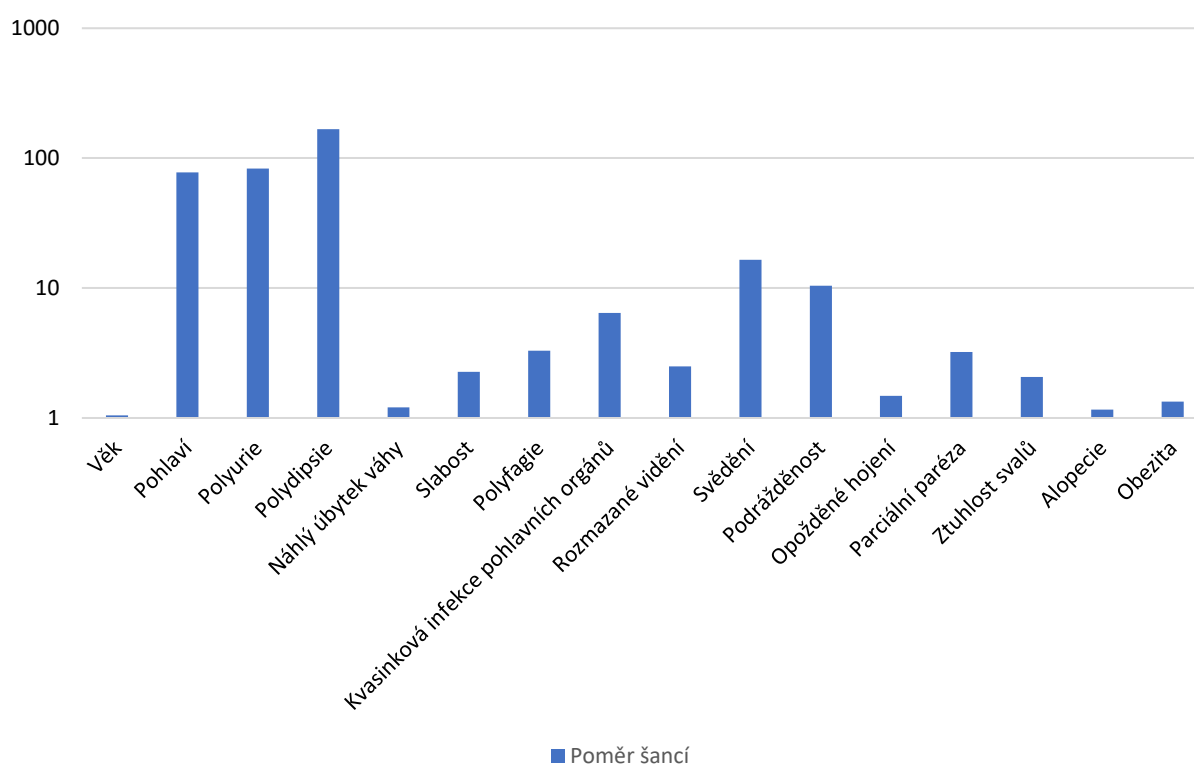
Logistická regrese byla počítána pomocí programu Statistica. Výsledky výpočtu jsou uvedeny v tabulce č. 1. Můžeme vidět, že většina proměnných se ukázala jako významných co do predikce diabetu, a to podle p-hodnot nižších než 0,05 a vysoké Waldovy statistiky. Jedná se konkrétně o faktory věk (vysoký), pohlaví (muž), přítomná polyurie, polydipsie, polyfagie, kvasinková infekce pohlavních orgánů, svědění, podrážděnost a parciální paréza. Jako nejsilnější faktory se ukázaly být přítomnost **polyurie** ( $p < 0,05$ ;  $W = 39,626$ ), **polydipsie** ( $p < 0,05$ ;  $W = 37,42$ ) a **mužské pohlaví** ( $p < 0,05$ ;  $W = 52,91$ ).

Vzhledem k nastavení referenčních kategorií v softwaru vyšly regresní koeficienty u hlavních symptomů se záporným znaménkem. Pro srozumitelnější interpretaci zvýšení šance na onemocnění byla použita reciproční hodnota původního poměru šancí ( $1/\text{Exp}(B)$ ). Po použití vzorce lze konstatovat, že přítomnost polyurie zvyšuje šanci na diabetes 167krát, přítomnost polydipsie 83krát a příslušnost k mužskému pohlaví 78krát (viz graf č.1).

Tabulka č.1 Regresní koeficienty a jejich testy statistické významnosti

	B	Směrodatná chyba	Waldova statistika	Stupně volnosti	p-hodnota	Exp(B)
<b>Věk</b>	0,051	0,025	4,07	1	<b>&lt;0,05</b>	1,053
<b>Pohlaví</b>	2,176	0,299	52,91	1	<b>&lt;0,05</b>	77,57
<b>Polyurie</b>	-4,44	0,705	39,626	1	<b>&lt;0,05</b>	0,012
<b>Polydipsie</b>	-5,07	0,829	37,42	1	<b>&lt;0,05</b>	0,006
Náhlý úbytek váhy	-0,19	0,548	0,121	1	0,728	0,827
Slabost	-0,817	0,537	2,317	1	0,128	0,442
<b>Polyfagie</b>	-1,194	0,534	5,007	1	<b>&lt;0,05</b>	0,303
<b>Kvasinková infekce pohlavních orgánů</b>	-1,864	0,553	11,345	1	<b>&lt;0,05</b>	0,155
Rozmazané vidění	-0,916	0,651	1,978	1	0,16	0,4
<b>Svědění</b>	2,803	0,673	17,362	1	<b>&lt;0,05</b>	16,493
<b>Podrážděnost</b>	-2,341	0,591	15,712	1	<b>&lt;0,05</b>	0,096
Opožděné hojení	0,392	0,55	0,507	1	0,476	1,479
<b>Parciální paréza</b>	-1,159	0,525	4,88	1	<b>&lt;0,05</b>	0,314
Ztuhlost svalů	0,728	0,58	1,578	1	0,209	2,073
Alopecie	-0,15	0,62	0,059	1	0,808	0,86
Obezita	0,289	0,544	0,282	1	0,595	1,335

Graf č. 1 Srovnání poměrů šancí u vybraných prediktorů diabetu



Po posouzení statistické věrohodnosti logistického modelu ( $X^2 = 521,275$ ;  $df = 16$ ;  $p < 0,05$ ) lze konstatovat, že model je statisticky významný. Tento výsledek potvrzuje, že zvolený soubor prediktorů společně významně přispívá k predikci diagnózy diabetu a model je jako celek funkční.

Kvalita sestaveného modelu byla posouzena pomocí koeficientů determinace. Hodnota Cox & Snell R-Square dosáhla 0,633, zatímco upravený Nagelkerke R-Square činil 0,86. Lze tedy interpretovat, že zvolené nezávislé proměnné vysvětlují přibližně **86 % variability** výskytu diabetu u sledovaného souboru respondentů. Zbývající část variability je způsobena faktory, které do modelu nebyly zahrnuty.

Tabulka č. 2 Ukazatelé kvality modelu

$X^2$	p-hodnota	Nagelkerke $R^2$	Cox & Snell $R^2$
521,275	<0,05	0,86	0,633

Pro ověření predikční schopnosti vytvořeného logistického modelu byla sestavena klasifikační matice (viz tabulka č. 3). Z výsledků vyplývá, že model dosahuje vysoké **celkové přesnosti 93,269 %**.

Tabulka č. 3 Přesnost odhadu modelu

Pozorované	Predikce		% správně
	Bez diabetu	Diabetes	
Bez diabetu	183	17	91,5 %
Diabetes	18	302	94,375 %
Celková přesnost			93,269 %

## ZÁVĚR

Tato práce měla za cíl ověřit pomocí logistické regrese, které symptomy typické pro diabetes jsou významnými a spolehlivými prediktory tohoto onemocnění. Z výsledků analýzy jednoznačně vyplývá, že nejsilnějšími varovnými signály jsou **polyurie** (nadměrné močení) a **polydipsie** (extrémní žízeň). U pacientů vykazujících tyto symptomy se šance na pozitivní diagnózu zvyšuje mnohonásobně oproti asymptomatickým jedincům. Mezi další významné faktory patří **pohlaví** (kde byla prokázána vyšší šance na výskyt u mužů).

Sestavený predikční model vykazuje **vysokou celkovou přesnost přesahující 93 %**, přičemž je velmi vyrovnaný v úspěšnosti identifikace nemocných i zdravých jedinců. 86% variability výskytu diabetu v souboru je vysvětlena právě kombinací těchto základních symptomů, což potvrzuje vysokou vypovídací hodnotu anamnestických dat.

V praxi tyto výsledky potvrzují, že jednoduchý screening zaměřený na specifické příznaky může sloužit jako efektivní nástroj pro včasné odhalení diabetu v běžné klinické praxi i v rámci prevence. Včasná identifikace rizikových pacientů na základě těchto symptomů je klíčová pro zahájení léčby a minimalizaci chronických komplikací spojených s tímto onemocněním.

## POUŽITÁ LITERATURA

American Diabetes Association. (2024). *Statistics about diabetes*. <https://diabetes.org/about-diabetes/statistics/about-diabetes>

Faniqul Islam, M. M., & Ferdousi, R., Sadikur, R., & Bushra, H. Y. (2020). Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques. *Computer Vision and*

*Machine Intelligence in Medical Image Analysis*, 922, 113–125. [https://doi.org/10.1007/978-981-13-8798-2\\_12](https://doi.org/10.1007/978-981-13-8798-2_12)

World Health Organization. (2023, April 5). *Diabetes*. <https://www.who.int/news-room/factsheets/detail/diabetes>