

Predikce diabetu mellitus typu 2 pomocí logistické regrese

Teoretické ukotvení

Diabetes mellitus je metabolická porucha charakterizovaná hyperglykemií vznikající v důsledku nedostatečné schopnosti organismu vylučovat inzulin nebo na něj správně reagovat (Khan et al., 2019). V současnosti patří mezi nejzávažnější onemocnění, a to zejména kvůli své vysoké prevalenci (Alex et al., 2022). Diabetes ovlivňuje nejen kvalitu života pacientů, ale představuje také velkou ekonomickou zátěž pro zdravotní systém (American Diabetes Association, 2010). S onemocněním jsou navíc spojeny další komplikace, například vysoký krevní tlak nebo onemocnění očí (Alex et al., 2022).

Dle studie Lugner et al. (2024) jsou hlavními prediktory vzniku diabetu mellitus typu 2 biologické a metabolické faktory, přičemž nejvýznamnější jsou hladina HbA1c, BMI, obvod pasu, poměr pas-boky, plazmatická glukóza, rodinná anamnéza diabetu, HDL cholesterol, věk a hladina sérového urátu, přičemž tyto faktory mají vyšší prediktivní sílu než životní styl, stravovací návyky nebo socioekonomický status.

Zpracování dat

Získaný soubor dat tvořily údaje 768 participantek, publikované Národním ústavem pro diabetes, onemocnění trávicí soustavy a ledvin. Do výzkumu byly zařazeny pouze ženy starší 21 let, které náleží do etnické skupiny Pima. Při kontrole souboru dat jsem našla poměrně vysoký počet chybějících údajů. Po očištění dat čítal soubor 393 participantek.

Pro zpracování dat jsem zvolila **binominální logistickou regresi**, která mi umožnila zkoumat vztahy mezi závislou proměnnou a několika nezávislými proměnnými. Pro tento účel jsem využila program Statistica.

Závislá proměnná je dichotomická, tedy nabývá pouze hodnot 1 nebo 0, přičemž v mém případě hodnota **1= přítomnost diabetu** a hodnota **0= nepřítomnost diabetu**.

Nezávislé proměnné byly následující:

- **Počet těhotenství**
- **Glukóza**
- **Krevní tlak**
- **Tloušťka kůže**

- Inzulín
- Věk

Výsledky logistické regrese

Tabulka č.1

Logistická regrese

	Odhad parametru	St. Chyba	Poměr šancí	Waldova Statistika	p
(Intercept)	-8,30	1,04		64,04	0,00
Počet těhotenství	0,05	0,05	1,05	0,92	0,34
Glukóza	0,04	0,01	1,04	44,71	0,00
Krevní tlak	0,00	0,01	1,00	0,14	0,71
Tloušťka kůže	0,04	0,01	1,04	9,99	0,00
Inzulín	0,00	0,00	1,00	0,01	0,94
Věk	0,03	0,02	1,04	3,70	0,05

Výsledky analýzy naznačují, že potenciálními prediktory rozvoje diabetu jsou nezávislé proměnné- **glukóza a tloušťka kůže**. Obě tyto proměnné jsou statisticky signifikantní a zároveň Waldova statistika u nich nabývá vysokých hodnot.

Tabulka č. 2

Kontrola kvality modelu

Ukazatele kvality modelu	
Cox-Snell R ²	Nagelkerke R ²
0,30	0,42

Koeficienty determinace dosáhly hodnot Cox-Snell R² = 0,30 a Nagelkerke R² = 0,42, což naznačuje, že model je schopen vysvětlit přibližně 30 až 42 % variability ve výskytu diabetu. Tyto hodnoty lze interpretovat jako střední míru vysvětlující schopnosti modelu.

Tabulka č. 3
Ukazatel predikce kvality modelu

Predikce diabetu	Predikce 0	Predikce 1	Procento správných
0	232	31	88,21
1	55	75	57,69

Tabulka č.3 ukazuje, že model správně zařadil 232 osob bez diabetu a 75 osob s diabetem. Přesnost predikce u osob bez onemocnění dosáhla 88,21 %, zatímco u osob s diabetem činila 57,69 %. Výsledky naznačují, že model je výrazně úspěšnější při identifikaci jedinců bez diabetu než při zachycení osob s onemocněním

Diskuse

Analýza ukázala, že zvýšená hladina glukózy a tloušťka kůže mohou být prediktory diabetu mellitus typu 2, což je v souladu s předchozími poznatky Lugner et al. (2024), kteří jako jeden z biologických a metabolických prediktorů identifikovali právě zvýšenou hladinu glukózy.

Hodnoty Cox-Snell $R^2 = 0,30$ a Nagelkerke $R^2 = 0,42$ spolu s výsledky klasifikace (88,21 % správně identifikovaných osob bez diabetu a 57,69 % osob s diabetem) naznačují střední schopnost modelu predikovat výskyt onemocnění.

Literatura

Alex, S. A., Nayahi, J. J. V., Shine, H., et al. (2022). Deep convolutional neural network for diabetes mellitus prediction. *Neural Computing & Applications*, 34, 1319–1327. <https://doi.org/10.1007/s00521-021-06431-7>

American Diabetes Association. (2010). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 33(Suppl. 1), S62–S69.

Khan, R. M. M., Chua, Z. J. Y., Tan, J. C., Yang, Y., Liao, Z., & Zhao, Y. (2019). From pre-diabetes to diabetes: Diagnosis, treatments and translational research. *Medicina*, 55(9), 546. <https://doi.org/10.3390/medicina55090546>

Lugner, M., Rawshani, A., Helleryd, E., et al. (2024). Identifying top ten predictors of type 2 diabetes through machine learning analysis of UK Biobank data. *Scientific Reports*, 14, 2102. <https://doi.org/10.1038/s41598-024-52023-5>

Data použita z: <https://www.kaggle.com/datasets/kandij/diabetes-dataset?resource=download>