

Profilácia UoZ pomocou strojového učenia

priebežná správa o analytickom projekte

december 2020

Martin Demeter

Tadeáš Chujac

Ján Komadel

Zhrnutie

Navrhujeme moderný prístup k profilácii uchádzačov o zamestnanie (UoZ), ktorý umožňuje v čase zaradenia osoby do evidencie UoZ posúdiť jej rizikovosť z hľadiska dlhodobého neuplatnenia sa na pracovnom trhu. Použitý model strojového učenia na základe historických údajov o rôznych uchádzačoch o zamestnanie kvantifikuje, nakoľko je pravdepodobné, že novo zaradený uchádzač si nenájde pracovné uplatnenie do 12 mesiacov. Test vykonaný na reálnych dátach z rokov 2015 až 2018 ukazuje, že strojové učenie dokáže u novo registrovaných UoZ pomerne spoľahlivo odhadnúť ich rizikovosť v tomto ohľade. Takéto vyhodnotenie potenciálu strojového učenia v oblasti práce s UoZ je dôležitým krokom v príprave na jej budúcu implementáciu do praxe.

Analýza potvrdzuje dôležitosť vzdelania a predchádzajúcej pracovnej aktivity pre zníženie rizika dlhodobého neuplatnenia sa. Zároveň aj ukazuje, že riziko neuplatnenia sa je pri čerstvých absolventoch škôl nižšie v porovnaní s inými uchádzačmi. Naopak, informácie o počítačových znalostiach vyplnené pri registrácii uchádzačov nenapomáhajú k lepšiemu odhadnutiu ich rizikovosti. Okrem všeobecného vplyvu jednotlivých charakteristík na rizikovosť dokáže použitá metóda identifikovať individuálne silné a slabé stránky z hľadiska nájdenia pracovného uplatnenia jednotlivých UoZ. Na základe našich doterajších zistení odporúčame realizovať systematickú včasnú profiláciu UoZ pomocou štatistických metód. Takýto prístup k profilácii umožňuje aj tvorbu samostatných štatistických modelov pre dôležité podskupiny UoZ (napr. absolventi alebo UoZ v pred dôchodkovom veku). Zároveň odporúčame zabezpečiť kontinuitu sledovania významných charakteristík UoZ, čím sa umožní aj aktualizácia štatistických modelov profilácie UoZ.

Obsah

1	Úvod.....	3
2	Použité údaje a metodika.....	6
3	Exploratívna analýza.....	7
4	Model LightGBM	10
5	Úspešnosť modelu	12
6	Prediktory rizikivosti	13
7	Modelové príklady určenia rizikivosti.....	15
8	Záver.....	17
	Literatúra.....	19
	Prílohy.....	20
A	Vysvetľujúce premenné.....	20
B	SHAP hodnoty pre ostatné prediktory.....	23
C	Modelové príklady nesprávne klasifikovaných uchádzačov	25
D	Znevýhodnenia a dĺžka evidencie	26

Abstrakt

Profilácia uchádzačov o zamestnanie (UoZ) s cieľom posúdenia rizikivosti osoby z hľadiska upadnutia do dlhodobej nezamestnanosti, prípadne dlhodobého neuplatnenia sa na pracovnom trhu, patrí vo viacerých krajinách k cenným nástrojom pri nastavení politiky trhu práce. Koncept dlhodobej nezamestnanosti rozširujeme na dlhodobé neuplatnenie sa na pracovnom trhu o prípady, kedy je UoZ vyradený z evidencie bez uplatnenia sa. Navrhujeme model strojového učenia, ktorý na základe údajov dostupných v čase zaradenia osoby do evidencie UoZ posúdi individuálnu rizikivosť toho, že tento UoZ do roka nenájde uplatnenie na trhu práce. Vybraný sofistikovaný model LightGBM dosahuje vyššiu úspešnosť ako častejšie využívané jednoduchšie modely typu logistickej regresie. Úspešná identifikácia rizikových uchádzačov okrem iného umožní lepšie cieľiť aktívne opatrenia trhu práce práve na tých, ktorí ich najviac potrebujú. Ako najsilnejšie charakteristiky zvyšujúce rizikivosť sa ukazujú absencia pracovného pomeru bezprostredne pred zaradením do evidencie, žiadne alebo len základné vzdelanie, prípadne dlhodobý predchádzajúci pracovný pomer s nízkou mzdou. Rizikovejší sú aj uchádzači bez vodičského preukazu. Naopak, absolventi sú vystavení nižšiemu riziku neuplatnenia sa. Počítačové znalosti vyplnené pri registrácii sa prekvapivo neukazujú ako významný prediktor. Pre konkrétnych UoZ identifikujeme pomocou SHAP hodnôt tie individuálne charakteristiky, ktoré najviac ovplyvňujú ich mieru rizikivosti nenájdenia si uplatnenia do 12 mesiacov od zaradenia do evidencie UoZ.

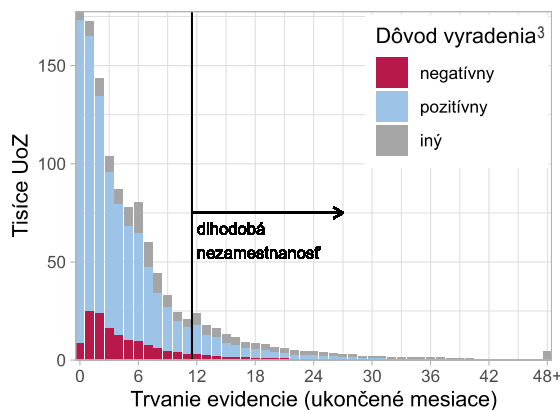
PodĎakovanie

Za konzultácie a cenné rady autori ďakujú Lucii Fašungovej, Štefanovi Domonkosovi a Ester Lujze Zatkalíkovej (Inštitút sociálnej politiky), Lukášovi Lafférsovi (Univerzita Mateja Bela), Martinovi Guzimu (Masarykova univerzita), Andrejovi Iringovi (Cognexa s.r.o.) a Vladimírovi Kvetanovi (Európske stredisko pre rozvoj odborného vzdelávania). Autori ďalej ďakujú kolegom z Ústredia práce, sociálnych vecí a rodiny, a to najmä za poskytnuté dáta a pomoc s ich interpretáciou a Sociálnej poisťovni za poskytnuté dáta.

1 Úvod

Viac ako štvrtina zaradení do evidencie uchádzačov o zamestnanie (UoZ) v rokoch 2015 až 2018 neskončila uplatnením sa uchádzača na trhu práce do jedného roka. Spomedzi 1,2 milióna prípadov nezamestnanosti zaradených do evidencie¹ v tomto období viac ako 175 tisíc (14,6 %) trvalo dlhšie ako rok a ďalších takmer 135 tisíc (11,2 %) bolo ukončených síce skôr ako po 12 mesiacoch, ale z dôvodov, ktoré nezodpovedajú uplatneniu sa na trhu práce (Obrázok 1, Tabuľka 1).² Hranica 12 mesiacov bola zvolená na základe slovenskej legislatívy (Box 1) a medzinárodne používaných definícií,³ podľa ktorých sa občan stáva dlhodobo nezamestnaným, ak jeho nezamestnanosť trvá aspoň jeden rok. Zo spoločenského hľadiska považujeme za nemenej problematickú aj skupinu tých UoZ, ktorí sa síce nestanú dlhodobo nezamestnanými, lebo sú z evidencie vyradení skôr, ale bez pracovného uplatnenia. V analýze sa preto nevenujeme iba dlhodobej nezamestnanosti ale dlhodobému neuplatneniu sa, ktoré v sebe navyše zahŕňa aj tieto prípady.

Obrázok 1 Každé siedme zaradenie do evidencie UoZ v rokoch 2015 až 2018 viedlo k dlhodobej nezamestnanosti



Zdroj: ÚPSVaR. Pozn.: Trvanie evidencie je zaokrúhlené na celé mesiace nadol. Posledná kategória „48+“ obsahuje evidencie trvajúce 48 mesiacov a viac.

Tabuľka 1 Takmer dve tretiny zaradení do evidencie UoZ v rokoch 2015 až 2018 sa skončili uplatnením sa uchádzača na trhu práce do jedného roka

Dôvod vyradenia ³	Dlhodobá nezamestnanosť		Spolu
	Nie	Áno	
negatívny	134 629	25 515	160 144
pozitívny	793 234	84 887	878 121
iný	97 149	64 868	162 017
spolu	1 025 012	175 270	1 200 282

Zdroj: ÚPSVaR Pozn.: Dlhodobá nezamestnanosť vzniká, keď je občan vedený v evidencii UoZ najmenej 12 po sebe nasledujúcich mesiacov.

Včasnú uplatnenie sa UoZ na trhu práce patrí medzi najzávažnejšie výzvy politiky trhu práce na Slovensku. Dlhodobá nezamestnanosť vedie k zhoršenej sociálnej situácii, strate pracovných návykov a zručností a často aj postupnej strate motivácie u nezamestnanej osoby (Soukup, 2011). Podobné nepriaznivé následky možno očakávať aj u osôb, ktoré sa nestanú dlhodobo nezamestnanými, lebo sú z evidencie vyradené skôr ako po roku, ale bez toho, aby si našli uplatnenie na trhu práce.

Počet dlhodobo nezamestnaných po nástupe koronakrízy vzrástol o viac ako 27 % (medzi februárom a októbrom 2020) a celkový počet uchádzačov o zamestnanie o takmer tretinu. Počas expanzie ekonomiky v posledných rokoch počet uchádzačov o zamestnanie klesal a s ním sa znižoval aj počet dlhodobo nezamestnaných (Obrázok 2). Vývoj počtu dlhodobo nezamestnaných bol ešte priaznivejší ako vývoj celkového počtu UoZ, keď ich podiel klesol z 55 % na začiatku roku 2015 na hodnoty okolo 35 % na prelome rokov 2019 a 2020 (Obrázok 3). Šírenie nového koronavírusu na území SR od marca 2020 a následné reštriktívne

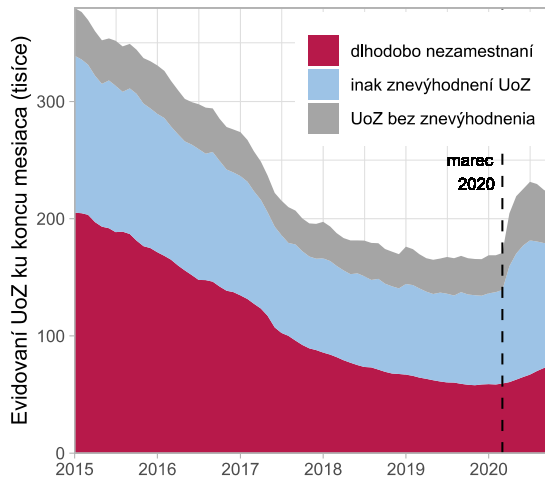
¹ Počet zaradení do evidencie nezodpovedá počtu osôb, nakoľko niektoré osoby boli počas daného obdobia zaradené do evidencie viackrát a v uvedených údajoch sú potom započítané viacnásobne. Jedinečných UoZ bolo v rokoch 2015 až 2018 zaradených do evidencie takmer 824 tisíc. Viaceré zaradenia rovnakej osoby nemožno považovať za duplicitné pozorovania, keďže mnohé dôležité premenné sa pri novej registrácii menia.

² Rozdelenie konkrétnych dôvodov vyradenia z evidencie UoZ na pozitívne (UoZ našiel uplatnenie na trhu práce), negatívne (UoZ nenašiel uplatnenie) a iné (dôvod vyradenia nedáva jasnú informáciu o uplatnení) obsahuje Tabuľka 3.

³ Napríklad definície dlhodobej nezamestnanosti OECD <https://data.oecd.org/unemp/long-term-unemployment-rate.htm> alebo Eurostatu https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Long-term_unemployment.

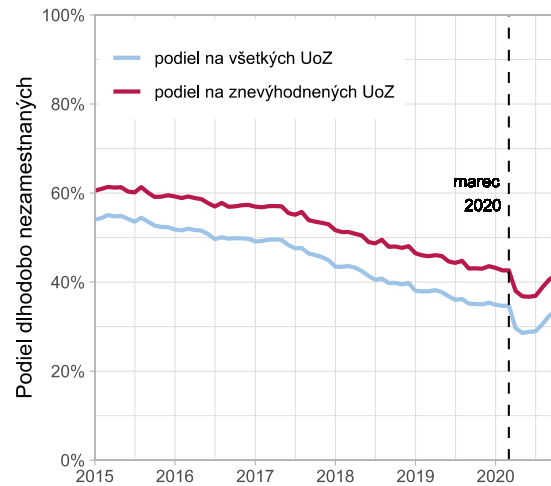
opatrenia však spôsobili na trhu práce šok, ktorý bol spojený s nárastom počtu UoZ aj počtu dlhodobo nezamestnaných.

Obrázok 2 Počet dlhodobo nezamestnaných aj celkový počet UoZ mal klesajúci trend až do koronakrízy



Zdroj: ÚPSVaR

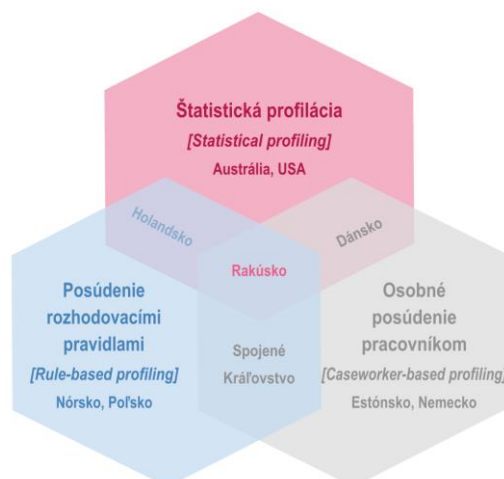
Obrázok 3 Podiel dlhodobo nezamestnaných UoZ v evidencii sa tiež do koronakrízy znižoval



Zdroj: ÚPSVaR

Najlepšia prax pri riešení dlhodobej nezamestnanosti je jej predchádzanie včasnou intervenciou v podobe aktívnych opatrení trhu práce. Profilácia UoZ s využitím údajov o uchádzačoch o zamestnanie dostupných už v čase ich zaradenia do evidencie umožňuje pomocou matematických a štatistických metód strojového učenia odhadnúť rizikovosť osoby z hľadiska upadnutia do dlhodobej nezamestnanosti, prípadne neuplatnenia sa na trhu práce do jedného roka, predtým, ako tento problém nastane. Štatistická profilácia UoZ na základe dát je moderným prístupom k profilácii, ktorá tradične prebieha buď na základe subjektívneho posúdenia pracovníka na úrade práce (*caseworker-based profiling*), alebo na základe rozhodovacích pravidiel napríklad podľa veku, pohlavia alebo vzdelania uchádzača (*rule-based profiling*). Obdobné moderné prístupy sa už niekoľko rokov používajú ako alternatíva alebo ako komplementárne nástroje k tradičnej profilácii UoZ vo viacerých štátoch OECD (Scoppetta & Buckenleib, 2018; Desiere, Langenbucher, & Struyven, 2019). V európskych krajinách je preferovaný skôr prístup kombinujúci štatistické modelovanie rizikovosti a posúdenie pracovníkmi, zatiaľ čo v Austrálii alebo v USA sa spoliehajú na výstupy modelov (Obrázok 4). Úspešná aplikácia profilácie okrem iného umožní včasne identifikovať rizikových UoZ a následne lepšie cieľiť aktívne opatrenia trhu práce práve na tých, ktorí ich najviac potrebujú.

Obrázok 4 Rôzne prístupy k profilácii UoZ v krajinách OECD



Zdroj: Desiere, Langenbucher, & Struyven (2019, s. 9)

Box 1 Znevýhodnení uchádzači o zamestnanie

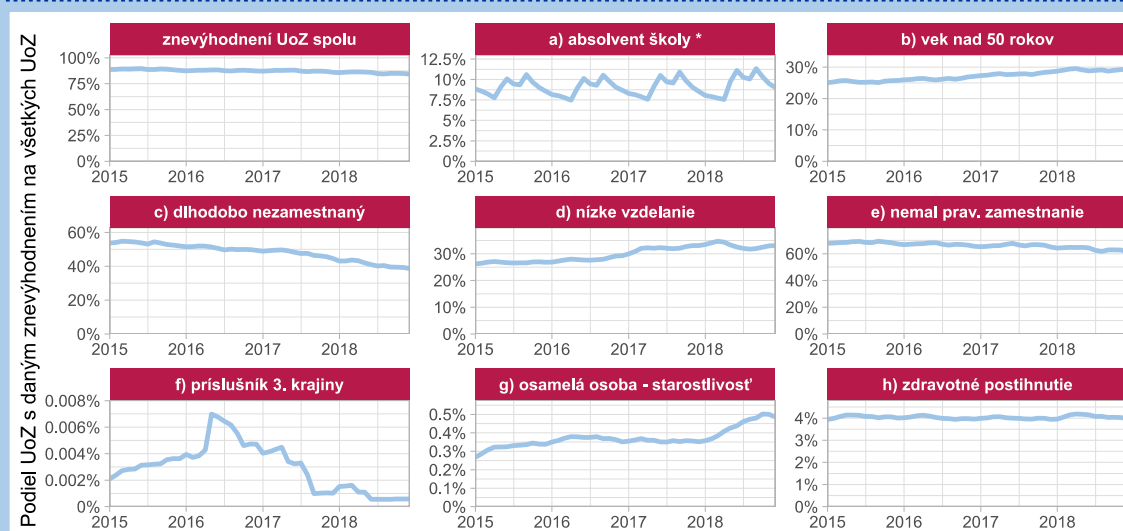
Viac ako osem z desiatich UoZ v evidencii má nejaké znevýhodnenie. Zákon o službách zamestnanosti definuje 8 druhov znevýhodnení,⁴ z ktorých najpočetnejšie sú znevýhodnenia hovoriace o nečinnosti na pracovnom trhu v poslednom roku pred zaradením do evidencie (Tabuľka 2). V priemere boli v rokoch 2015 až 2018 takmer dve tretiny UoZ bez pravidelného plateného zamestnania⁵ v posledných 12 mesiacoch pred zaradením do evidencie a takmer polovica UoZ bola dlhodobo nezamestnaná. Časté znevýhodnenia sú tiež nízke vzdelanie (32 % UoZ) a vek nad 50 rokov (27 % UoZ). Naopak, najmenej zastúpení sú štátni príslušníci tretích krajín s azylom alebo doplnkovou ochranou, ktorých bolo najviac len 15 (0,005 % UoZ v júni 2016), a osamelé osoby starajúce sa o inú osobu (0,5% UoZ).

Tabuľka 2 Najpočetnejšie znevýhodnenia medzi UoZ, ktorí boli v evidencii niekedy v rokoch 2015 až 2018, sú absencia pravidelného zamestnania za posledný rok a dlhodobá nezamestnanosť

Znevýhodnenie	Priemerný počet	Priemerný podiel
a) absolvent školy - občan mladší ako 26 rokov, ktorý ukončil sústavnú prípravu na povolanie v dennej forme štúdia pred menej ako dvomi rokmi a od jej ukončenia nemal pravidelne platené zamestnanie	14 965	5,51%
b) občan starší ako 50 rokov	71 602	27,27%
c) dlhodobo nezamestnaný - občan vedený v evidencii UoZ najmenej 12 po sebe nasledujúcich mesiacov	132 092	48,46%
d) občan, ktorý dosiahol vzdelanie nižšie ako stredné odborné vzdelanie	84 296	31,91%
e) občan, ktorý najmenej 12 po sebe nasledujúcich kalendárnych mesiacov pred zaradením do evidencie UoZ nemal pravidelne platené zamestnanie	177 242	66,16%
f) štátny príslušník tretej krajiny, ktorému bol udelený azyl alebo poskytnutá doplnková ochrana	7	0,00%
g) občan, ktorý žije ako osamelá dospelá osoba s aspoň jednou osobou odkázanou na jeho starostlivosť alebo starajúca sa aspoň o jedno dieťa pred skončením povinnej školskej dochádzky	1 375	0,51%
h) občan so zdravotným postihnutím	9 544	3,57%
UoZ s aspoň jedným znevýhodnením	231 936	86,63%
všetci evidovaní UoZ	266 204	

Zdroj: ÚPSVaR. Pozn.: Uvedené sú priemery za mesiace január 2015 až december 2018.

Obrázok 5 V rokoch 2015 až 2018 nastal výrazný pokles len pri podiele dlhodobo nezamestnaných



Zdroj: ÚPSVaR Pozn.: *Pri absolventoch školy uvádzame UoZ, ktorí mali toto znevýhodnenie v čase zaradenia do evidencie, keďže toto znevýhodnenie platí len dva roky od ukončenia sústavnej prípravy na povolanie. Pri ostatných znevýhodneniach uvádzame UoZ, ktorí dané znevýhodnenie mali ku koncu zobrazeného mesiaca.

⁴ Znevýhodnených uchádzačov o zamestnanie definuje § 8 zákona č. 5/2004 Z. z. o službách zamestnanosti.

⁵ Zamestnanie, ktoré trvalo najmenej šesť po sebe nasledujúcich mesiacov podľa § 8 zákona č. 5/2004 Z. z. o službách zamestnanosti.

Podiel znevýhodnených UoZ na všetkých evidovaných UoZ v rokoch 2015 až 2018 klesol z takmer 90 % na 84 % (Obrázok 5). Klesajúci trend pozorujeme aj pri najčastejších znevýhodneniach (občania bez pravidelného zamestnania a dlhodobo nezamestnaní). Na druhej strane, podiel UoZ nad 50 rokov v rovnakom období vzrástol z 25 % na takmer 30 % a podiel UoZ s nízkym vzdelaním sa zvýšil z 26 % na 33 %. Pri absolventoch dochádza každoročne k prirodzenému nárastu v júni a v septembri, ale celkový trend je len mierne rastúci. Podrobnejšia štruktúra znevýhodnených UoZ podľa dĺžky evidencie je v prílohe D.

2 Použité údaje a metodika

Základným zdrojom údajov pri profilácii uchádzačov o zamestnanie je evidencia UoZ, ktorú vedie Ústredie práce, sociálnych vecí a rodiny (ÚPSVaR). Pri zostavení modelu, validácii aj testovaní úspešnosti pracujeme s dátami o uchádzačoch, ktorí boli zaradení do evidencie v rokoch 2015 až 2018.⁶

Vysvetľovanou premennou je binárna premenná y , ktorá **identifikuje uchádzačov, ktorí nenašli uplatnenie na trhu práce do jedného roka** od zaradenia do evidencie. Konkrétne definujeme $y = 1$, pre tých UoZ, ktorí buď (i) neboli vyradení z evidencie skôr ako jeden rok po zaradení (stali sa dlhodobo nezamestnanými), alebo (ii) boli vyradení z evidencie skôr ako jeden rok po zaradení, ale z dôvodov, ktoré nezodpovedajú uplatneniu sa na trhu práce (negatívne dôvody v Tabuľke 3). Pre uchádzačov, ktorí našli uplatnenie na trhu práce skôr ako 1 rok po zaradení (pozitívne dôvody), definujeme $y = 0$. Pochopiteľne, existujú aj dôvody vyradenia z evidencie, pri ktorých nie je možné vo všeobecnosti určiť, či sú z hľadiska uplatnenia sa UoZ na trhu práce pozitívne alebo negatívne. Tieto dôvody uvádzame v poslednej časti Tabuľky 3 a UoZ vyradených z evidencie skôr ako po roku z týchto dôvodov do modelov nezahŕňame.

Tabuľka 3 Viac ako polovica zaradení do evidencie UoZ v rokoch 2015 až 2018 skončila vznikom pracovného vzťahu

Dôvod vyradenia podľa § 36	Kód	Popis	Počet výskytov
Pozitívne dôvody - uplatnenie na trhu práce			878 121
ods. 1 písm. a)	V01	vznik pracovného pomeru alebo obdobného pracovného vzťahu	624 682
ods. 1 písm. b)	V02	vznik oprávnenia prevádzkovať alebo vykonávať SZČ	60 757
ods. 1 písm. c)	V03	skončenie pozastavenia prevádzkovania alebo pozastavenia vykonávania SZČ	5 260
ods. 1 písm. d)	V04	nástup na sústavnú prípravu na povolanie	14 269
ods. 1 písm. j)	V10	odchod do členského štátu EÚ na obdobie dlhšie ako 15 kalendárnych dní	40 326
ods. 1 písm. k)	V11	odchod do cudziny na obdobie dlhšie ako 15 kalendárnych dní	11 651
ods. 1 písm. l)	V12	začatie vykonávania zárobkovej činnosti v EÚ alebo v cudzine	73 468
ods. 1 písm. o)	V15	vznik PPV (dohody mimo PP) alebo podľa osobitného predpisu	1 059
ods. 1 písm. p)	V16	nadobudnutie právoplatnosti rozsudku o neplatnosti skončenia zamestnania	42
ods. 1 písm. s)	V18	prestal splňať podmienku podľa § 6 ods. 2 písm. a)	21 789
ods. 1 písm. s)	V20	prestal splňať podmienku podľa § 6 ods. 2 písm. a) a b) a ods. 3.	24 818
Negatívne dôvody - neuplatnenie na trhu práce			160 144
ods. 1 písm. e)	V05	nástup na výkon trestu odňatia slobody	1 073
ods. 1 písm. f)	V06	vzatie do výkonu väzby	1 071
ods. 2 písm. a), b), c)	V19	vykonávanie nelegálnej práce, nespolupráca s úradom alebo udelené povolenie na zamestnanie v cudzine	158 000
Iné dôvody - nejasné uplatnenie na trhu práce			162 017
ods. 1 písm. g)	V07	priznanie starobného dôchodku alebo dovŕšenie veku	6 204
ods. 1 písm. h)	V08	narodenie dieťaťa	24 804
ods. 1 písm. i)	V09	úmrtie UoZ	2 604
ods. 1 písm. m)	V13	požiadanie o vyradenie z dôvodu starostlivosti o dieťa do 10 rokov alebo o blízkou osobu	15 039
ods. 1 písm. n)	V14	požiadanie o vyradenie (vlastná žiadosť)	77 609
ods. 1 písm. r)	V17	nesprávne zaradenie do evidencie UoZ	61
ods. 1 písm. t)	V21	zánik dlhodobého pobytu štátneho príslušníka tretej krajiny	17
ods. 1 písm. u)	V23	neschopnosť plniť povinnosti UoZ podľa posudku lekára	193
		neurčený dôvod vyradenia	28
		ešte nevyradení	35 458

Zdroj: ÚPSVaR. Pozn.: Uvedené sú dôvody vyradenia z evidencie UoZ podľa § 36 zákona o službách zamestnanosti.

⁶ Pri novších údajoch na základe v súčasnosti dostupných údajov nie je známe, či nezamestnanosť trvala viac ako rok.

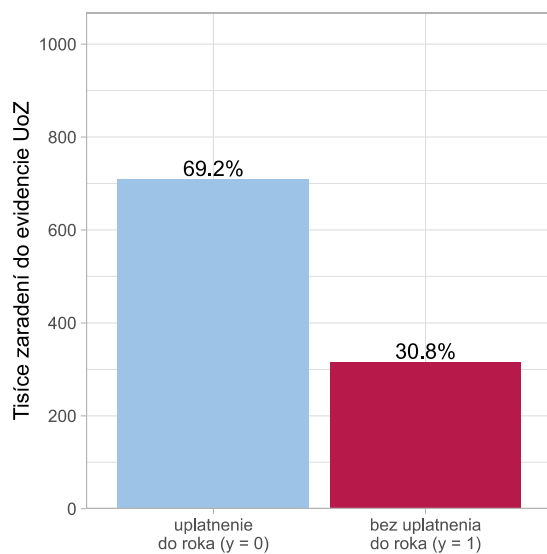
Výsledný dátový súbor obsahuje 1,026 milióna pozorovaní a 90 vysvetľujúcich premenných, ktoré sú bližšie opísané v prílohe A.⁷ Vysvetľovanú premennú y modelujeme pomocou charakteristík daného uchádzača evidovaných pri zaradení do evidencie UoZ, doplnených o údaje o v minulosti absolvovaných aktívnych opatreniach trhu práce (AOTP), o miere nezamestnanosti a evidovaných voľných pracovných miestach v okrese trvalého pobytu UoZ v čase registrácie, o poberaní dávok v hmotnej núdzi, o stupni odkázanosti a miere funkčnej poruchy, o predchádzajúcej ekonomickej aktivite a predchádzajúcich príjmoch UoZ na základe poistných vzťahov Sociálnej poisťovne a o poberaní dôchodku.⁸

3 Exploratívna analýza

V rôznych skupinách populácie UoZ pozorujeme rozdielne zastúpenie uchádzačov, ktorí nenašli uplatnenie do jedného roka. Exploratívna analýza pomáha odhaliť medzi vstupnými premennými potenciálne prediktory vysvetľovanej premennej y .

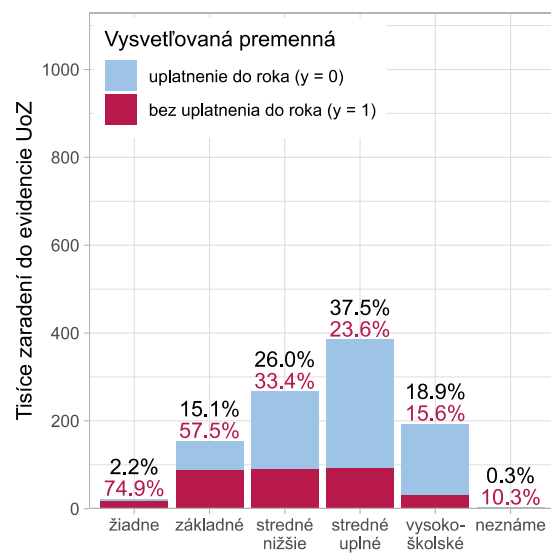
Skoro tretina zaradení do evidencie UoZ skončí neuplatnením sa na pracovnom trhu do 12 mesiacov. Rozdelenie modelovanej premennej y vo finálnom dátovom súbore je mierne nevyvážené. Viac ako 69 % z celkového počtu evidencií tvoria UoZ, ktorým sa podarilo nájsť si uplatnenie na trhu práce do jedného roka od zaradenia do evidencie (Obrázok 6). Zvyšných takmer 31 % zaradení do evidencie UoZ skončilo dlhodobou nezamestnanosťou alebo negatívnym vyradením z evidencie UoZ.

Obrázok 6 Traja z desiatich UoZ vo výslednom dátovom súbore nenašli uplatnenie do jedného roka



Zdroj: ÚPSVaR

Obrázok 7 S vyšším dosiahnutým vzdelaním klesá podiel UoZ, ktorí nenájdu uplatnenie



Zdroj: ÚPSVaR Pozn.: Čiernou je uvedené zastúpenie skupiny v populácii UoZ, červenou zastúpenie rizikových UoZ v skupine.

Takmer 60% UoZ so žiadnym alebo základným vzdelaním zaradených do evidencie nenašlo uplatnenie na trhu práce (Obrázok 7). Skupiny UoZ s najnižším vzdelaním sú jednoznačne najrizikovejšie z pohľadu možnosti upadnutia do dlhodobej nezamestnanosti. Pravdepodobnosť výskytu osôb s dlhodobým problémom uplatnenia sa v jednotlivých častiach populácie rozdelenej na základe najvyššieho dosiahnutého vzdelania úmerne klesá s nárastom stupňa dosiahnutého vzdelania. Menej ako 16 % UoZ, ktorí dokončili nejaký stupeň vysokoškolského štúdia, nie je schopných nájsť si uplatnenie na trhu práce.

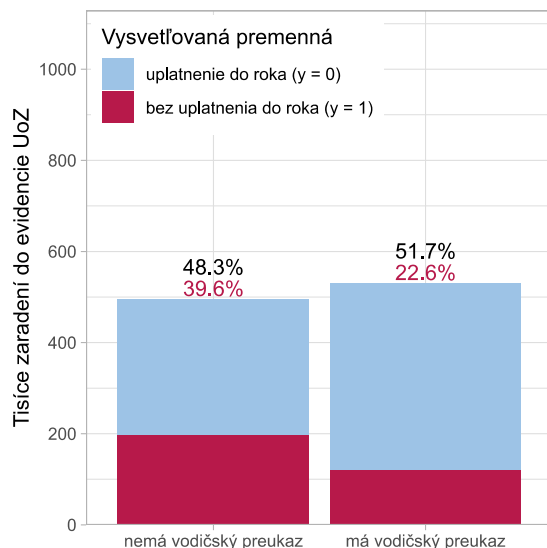
⁷ Z pôvodného počtu 1,2 milióna zaradení do evidencie boli zlúčené do jedného pozorovania po sebe nasledujúce zaradenia rovnakej osoby, medzi ktorými prešlo menej ako dva mesiace, a boli odstránené zaradenia, ktoré skončili skôr ako po 12 mesiacoch dôvodom vyradenia s nejasným uplatnením na trhu práce. Navyše bolo odstránených aj asi 100 pozorovaní, pri ktorých chýbal údaj o trvalom pobyte alebo o veku UoZ.

⁸ Na spracovanie údajov aj tvorbu obrázkov a sme využili prostredie R (R Core Team, 2019) s knižnicami obsiahnutými v metabalíku *tidyverse* zameranom na dátovú vedu (Wickham, a iní, 2019).

UoZ disponujúci základným vodičským preukazom majú vyššiu pravdepodobnosť nájsť si prácu do jedného roka od zaradenia do evidencie.⁹ Podiel uchádzačov so základným vodičským preukazom, ktorým sa to nepodarilo, tvorí menej ako 23 %. Naopak, zvýšenú koncentráciu (takmer 40 %) uchádzačov nenachádzajúcich uplatnenie je možné pozorovať pri UoZ, ktorí základným vodičským preukazom nedisponujú (Obrázok 8).

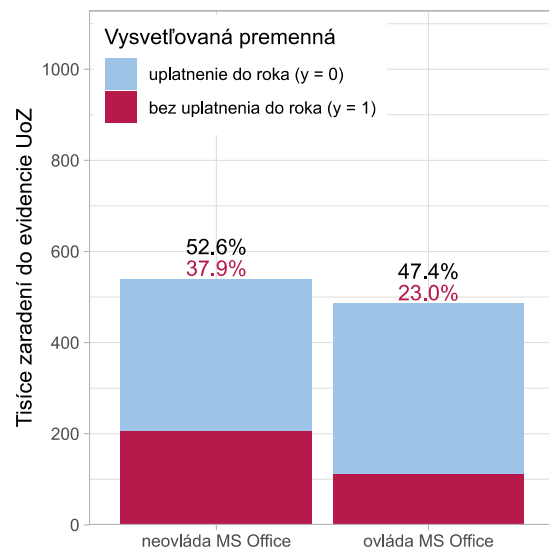
Prvotná deskriptívna analýza naznačuje, že počítačové znalosti sú dôležité pri hľadaní práce. Spomedzi uchádzačov, ktorí v žiadosti o zaradenie do evidencie UoZ uviedli schopnosť používať základné nástroje MS Office na základnej alebo pokročilej úrovni, 23 % nenašlo do roka uplatnenie na trhu práce.¹⁰ Naopak hustota zastúpenia UoZ bez uplatnenia je nepomerne vyššia medzi uchádzačmi, ktorí tieto nástroje neovládajú, a dosahuje až takmer 38 % (Obrázok 9). Informácie o počítačových znalostiach vyplnené pri registrácii uchádzačov sa napriek tomu neukázali byť významné v kontexte výsledného modelu strojového učenia a nevedli k lepšiemu odhadu ich rizikovosti.

Obrázok 8 Medzi UoZ bez vodičského preukazu je zastúpenie UoZ bez uplatnenia o tri štvrtiny vyššie



Zdroj: ÚPSVaR. Pozn.: Čiernou je uvedené zastúpenie skupiny v populácii UoZ, červenou zastúpenie rizikových UoZ v skupine. Zohľadnené sú základné vodičské preukazy skupín A, A1, A2, AM, B a B1.

Obrázok 9 Medzi UoZ, ktorí neovládajú MS Office, je zastúpenie UoZ bez uplatnenia o dve tretiny vyššie



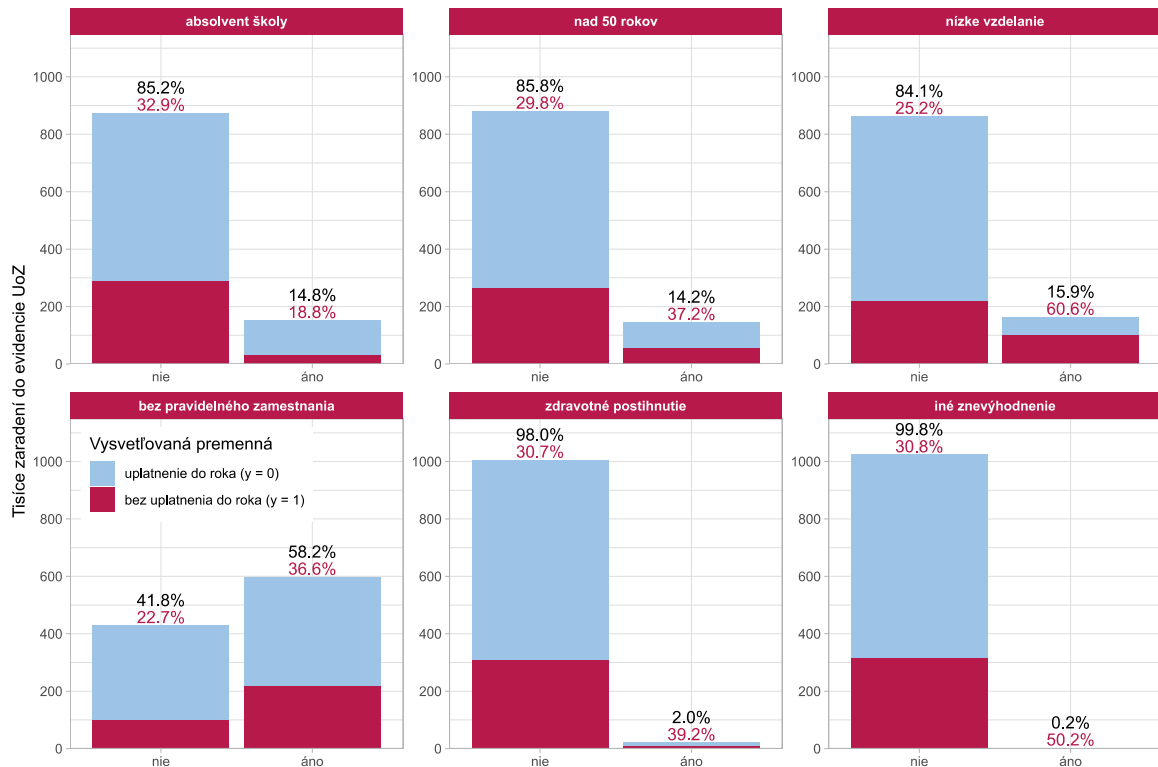
Zdroj: ÚPSVaR. Pozn.: Čiernou je uvedené zastúpenie skupiny v populácii UoZ, červenou zastúpenie rizikových UoZ v skupine. Zohľadnené je, či UoZ uviedol, že ovláda základné nástroje MS Office na základnej alebo pokročilej úrovni.

Absolventi škôl sú jedinou znevýhodnenou skupinou UoZ, pri ktorej je zastúpenie UoZ bez uplatnenia sa na trhu práce do jedného roka dokonca nižšie ako pri tých, ktorí žiadne znevýhodnenie nemajú. Iba 18,8 % z absolventov škôl, ktorých na základe Zákona o službách zamestnanosti zaraďujeme medzi znevýhodnených UoZ (Box 1), upadá do dlhodobej nezamestnanosti alebo je vyradených z evidencie UoZ z negatívneho dôvodu (Obrázok 10). Zvyšné skupiny znevýhodnených UoZ vykazujú očakávané hodnoty z pohľadu koncentrácie neúspešných uchádzačov, kedy znevýhodnenie naozaj poukazuje na zvýšené zastúpenie UoZ s dlhodobým problémom s uplatnením sa na trhu práce. Avšak, na základe analýzy premennej označujúcej celkový počet znevýhodnení UoZ môžeme pozorovať, že napriek pomerne dobrému uplatneniu absolventov škôl na trhu práce je dlhodobým neuplatnením sa najmenej ohrozenou skupina UoZ bez akéhokoľvek znevýhodnenia (Obrázok 11).

⁹ Ako základné vodičské preukazy berieme skupiny A, A1, A2, AM, B a B1 (príloha A).

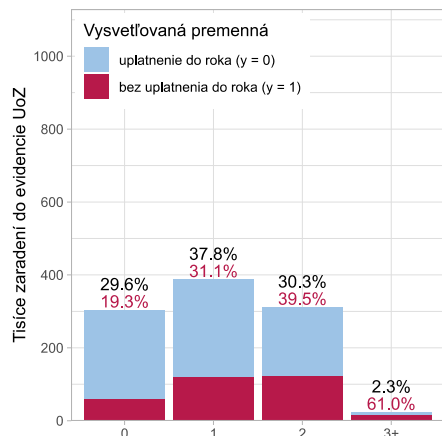
¹⁰ Za základné nástroje MS Office považujeme Word, Excel, PowerPoint a Outlook. Ovládanie týchto nástrojov na vysokej úrovni a ovládanie iných nástrojov MS Office zohľadňujeme v iných premenných (príloha A).

Obrázok 10 Absolvent školy je jediné znevýhodnenie, pri ktorom pozorujeme podpriemerné zastúpenie UoZ bez uplatnenia



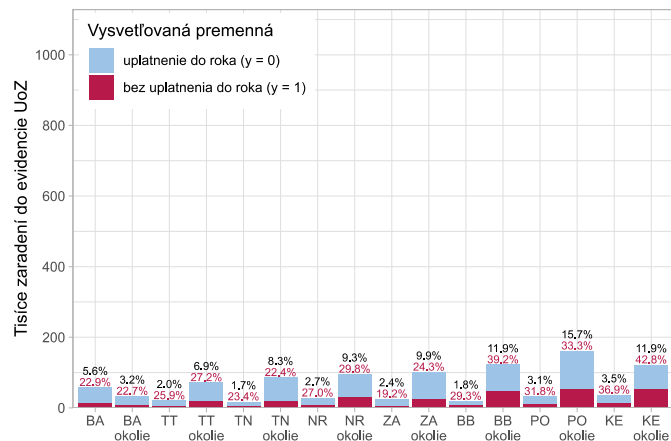
Zdroj: ÚPSVaR. Pozn.: Čiernou je uvedené zastúpenie skupiny v populácii UoZ, červenou zastúpenie rizikových UoZ v skupine.

Obrázok 11 Podiel UoZ bez uplatnenia rastie s počtom znevýhodnení¹¹



Zdroj: ÚPSVaR. Pozn.: Čiernou je uvedené zastúpenie skupiny v populácii UoZ, červenou zastúpenie rizikových UoZ v skupine.

Obrázok 12 Na východe je vyšší podiel uchádzačov, ktorí do roka nenachádzajú uplatnenie



Zdroj: ÚPSVaR. Pozn.: Čiernou je uvedené zastúpenie skupiny v populácii UoZ, červenou zastúpenie rizikových UoZ v skupine. Každý z krajov delíme na okresy krajského mesta a ostatné okresy v kraji.

Z regionálneho hľadiska sú uchádzači s dlhodobým problémom uplatnenia sa viac koncentrovaní na východe Slovenska. Najvyššie zastúpenie majú medzi UoZ s trvalým pobytom v Košickom kraji (viac ako 41 %) a najnižšie v Trenčianskom, Bratislavskom a Žilinskom kraji (okolo 23 %). Spomedzi okresov krajských

¹¹ Obrázok ukazuje, že 70,4 % zaradení do evidencie tvoria UoZ s aspoň jedným znevýhodnením. V Boxe 1 uvádzame, že v sledovanom období bolo znevýhodnených 78 % až 90 % UoZ, ktorí sa v daných mesiacoch nachádzali v evidencii. Nižší podiel znevýhodnených UoZ na zaradeniach do evidencie ako na stave evidencie v konkrétnom čase je spôsobený tým, že títo uchádzači spravidla ostávajú v evidencii dlhšie ako UoZ bez znevýhodnenia.

miest je najnižšia koncentrácia uchádzačov, ktorí si do roka nenašli uplatnenie, v Žiline (19,2 %). Naopak, v Košickom kraji okrem Košíc je ich podiel až 42,8 % (Obrázok 12).

Prieskumná analýza finálneho dátového súboru odhalila prostredníctvom pozorovaní podielu UoZ s dlhodobým problémom s pracovným uplatnením v rôznych častiach populácie UoZ viaceré nádejné prediktory vysvetľovanej premennej y . Úroveň najvyššie dosiahnutého vzdelania, vodičský preukaz, počítačové znalosti, počet znevýhodnení alebo miesto trvalého bydliska majú potenciál byť dobrými ukazovateľmi rizikosti neuplatnenia sa UoZ na trhu práce do jedného roka.

4 Model LightGBM

Na posúdenie miery rizikosti, že konkrétny novo zaradený UoZ nenájde uplatnenie na trhu práce do jedného roka od zaradenia do evidencie UoZ, používame techniku strojového učenia *gradient boosting* implementovanú v knižnici LightGBM. Pre jednoduchosť použitý model označujeme ako **model LightGBM**. Jeho vstupom sú vysvetľujúce premenné x (charakteristiky UoZ, Príloha A) a výstupom je pravdepodobnosť $p(x)$, že UoZ s týmito charakteristikami nenájde uplatnenie do jedného roka. Na tréningovej vzorke uchádzačov¹² sú odhadnuté parametre modelu tak, aby sa minimalizovala účelová funkcia, ktorou je logistická strata (*log loss*)

$$L_p(X, y) = -\frac{1}{n} \sum_{i=1}^n (\omega y_i \ln p(x_i) + (1 - y_i) \ln[1 - p(x_i)]),$$

kde x_i je vektor charakteristík i -teho UoZ, X je matica so stĺpcami x_i , vysvetľovaná binárna premenná y_i indikuje, či i -ty UoZ do roka našiel uplatnenie, a n je počet UoZ v tréningovej vzorke. Parameter ω umožňuje zohľadniť nevyváženosť zastúpenia pozitívnej ($y = 1$) a negatívnej triedy ($y = 0$).¹³

LightGBM sa radí do skupiny GBDT (*gradient boosting decision tree*) algoritmov, ktoré v sebe kombinujú viacero rozhodovacích stromov.¹⁴ Na minimalizáciu účelovej funkcie $L_p(X, y)$ využívajú postup nazývaný *gradient boosting*, pri ktorom je každý ďalší strom konštruovaný tak, aby jeho pridaním do modelu poklesla hodnota účelovej funkcie čo najviac (Hastie, Tibshirani, & Friedman, 2009). Výhodou LightGBM oproti iným GBDT metódam je jeho rýchlosť, efektivita a menšie nároky na pamäť počítača, ktoré dosahuje vďaka inovatívnemu GOSS (*Gradient-based One-Side Sampling*) algoritmu (Ke, a iní, 2017).

Model LightGBM dosiahol najlepšie výsledky z hľadiska ROC AUC na validačnej vzorke spomedzi šiestich testovaných metód strojového učenia (Obrázok 13, Box 2). Aj keď rozdiely v AUC nie sú veľké, v prospech modelov založených na rozhodovacích stromoch (LightGBM, XGBoost, náhodný les) oproti jednoduchej logistickej regresii hovorí ich schopnosť prirodzene zachytiť interakcie medzi premennými (Hastie, Tibshirani, & Friedman, 2009). GBDT modely navyše na rozdiel od náhodného lesu alebo logistickej regresie nemajú problém ani v prípade korelovaných premenných (Chen, He, Benesty, & Tang, 2020). Už spomínaný inovatívny GOSS algoritmus zvyhodňuje LightGBM oproti modelu XGBoost z hľadiska rýchlosti, pričom čas potrebný na natréňovanie modelu je menej ako tretinový. Vo zvolenom modeli LightGBM sme pre zjednodušenie vybrali len

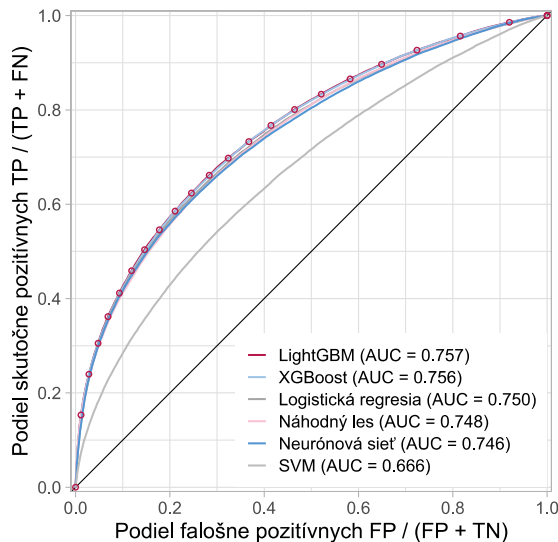
¹² Dátový súbor opísaný v časti 2 delíme na tréningovú (80 % pozorovaní) a testovaciu vzorku (20 %) so zachovaním zastúpenia vysvetľovanej premennej y . Tréningová vzorka je použitá na odhadnutie parametrov modelu a neskôr je testovacia vzorka použitá na *out-of-sample* hodnotenie úspešnosti. Na ladenie hyperparametrov algoritmu používame 5-násobnú krížovú validáciu na tréningovej vzorke, pri ktorej sme tiež zachovali zastúpenie hodnôt premennej y .

¹³ V našom prípade zohľadňujeme zhruba polovičné zastúpenie pozitívnej triedy v porovnaní s negatívnou triedou použitím $\omega = 2$, čím dostáva chyba pri minoritnej (pozitívnej) triede dvojnásobnú váhu.

¹⁴ Rozhodovacie stromy (v našom prípade klasifikačné) tu slúžia ako tzv. *base learner*.

40 najdôležitejších premenných podľa SHAP hodnôt¹⁵ a následnou ďalšou optimalizáciou hyperparametrov sme sa dopracovali k finálnemu preferovanému modelu, ktorého AUC dosahuje hodnotu 0,784 (Obrázok 14).¹⁶

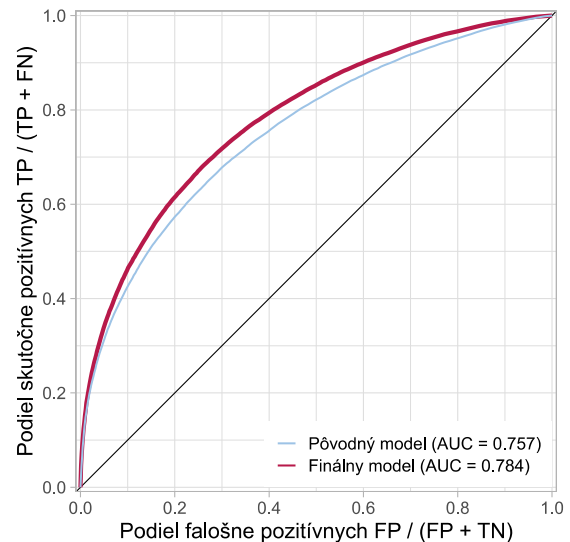
Obrázok 13 LightGBM dosiahol najvyššiu hodnotu ROC AUC spomedzi testovaných modelov



Zdroj: vlastné spracovanie

Pozn.: TP je počet prípadov, ktoré sú pozitívne ($y = 1$) a aj klasifikované ako rizikové ($\hat{y} = 1$); TN je počet prípadov, ktoré sú negatívne ($y = 0$) a aj klasifikované ako nerizikové ($\hat{y} = 0$); FP je počet prípadov, ktoré sú negatívne, ale klasifikované sú ako rizikové; FN je počet prípadov, ktoré sú pozitívne, ale klasifikované sú ako nerizikové.

Obrázok 14 Dodatočným ladením po redukcii počtu premenných sa kvalita modelu zvýšila



Zdroj: vlastné spracovanie

Box 2 Miery kvality modelu

Pri klasifikačnom modeli je prirodzené sledovať **správnosť klasifikácie** (*accuracy*), teda akú časť UoZ klasifikuje model správne – tých, ktorí do roka našli uplatnenie, ako nerizikových a tých, ktorí uplatnenie nenašli, ako rizikových (Tabuľka 4). V prípade nevyváženého dátového súboru, kde je niektorá z kategórií zastúpená výrazne viac ako iná, však **správnosť klasifikácie** môže poskytovať skreslený obraz o kvalite modelu.

Presnosť (*precision*) meria, aká časť UoZ, ktorí sú modelom označení ako rizikovní, skutočne nenájde do roka uplatnenie. **Citlivosť** (*sensitivity*) hovorí o tom, akú časť UoZ, ktorí skutočne do roka nenájdu uplatnenie, model označí ako rizikových. Naopak, **špecificita** (*specificity*) prezrádza, aká časť UoZ, ktorí skutočne nájdu do roka uplatnenie, je označená za nerizikových. Možný nepomer medzi početnosťami kategórií zohľadňuje aj **vyvážená správnosť klasifikácie** (*balanced accuracy*), ktorá priemeruje správnosť spočítanú pre každú z kategórií zvlášť.

F_1 štatistika sa definuje ako harmonický priemer presnosti a citlivosti. Na dosiahnutie vysokej hodnoty F_1 štatistiky preto musí mať model vysokú presnosť (veľká časť UoZ označených ako rizikovní skutočne nenájde uplatnenie) aj vysokú citlivosť (veľká časť UoZ, ktorí nenájdu uplatnenie, je označená ako rizikovní).

F_β štatistika zovšeobecňuje tento koncept a umožňuje priradiť rôznu dôležitosť presnosti a citlivosti. Parameter $\beta > 0$ vyjadruje, koľkokrát je citlivosť dôležitejšia ako presnosť. Napríklad F_2 štatistika dáva dvojnásobnú váhu citlivosti oproti presnosti, kým $F_{0,5}$ štatistika dáva citlivosti polovičnú váhu.

¹⁵ Dôležitosť premenných založená na SHAP hodnotách je na rozdiel od iných mier dôležitosti jediná konzistentná aj presná (Lundberg & Lee, 2017). 40 najdôležitejších premenných bolo zvolených preto, že pri tomto počte ešte neprišlo k poklesu kvality modelu a zároveň sa počet premenných znížil o viac ako polovicu.

¹⁶ Zoznam vybraných premenných obsahuje Tabuľka 6 v Prílohe A.

Tabuľka 4 Vybrané miery úspešnosti modelu binárnej klasifikácie

Miera úspešnosti	Anglický názov	Definícia ^a
správnosť	<i>accuracy</i>	$\frac{TP + TN}{TP + TN + FP + FN}$
presnosť	<i>precision, positive predictive value (PPV)</i>	$\frac{TP}{TP + FP}$
citlivosť ^b	<i>sensitivity, recall, true positive rate (TPR)</i>	$\frac{TP}{TP + FN}$
špecificita ^c	<i>specificity, selectivity, true negative rate (TNR)</i>	$\frac{TN}{TN + FP}$
vyvážená správnosť	<i>balanced accuracy</i>	$\frac{\text{citlivosť} + \text{špecificita}}{2}$
F_1 štatistika	F_1 score	$\frac{2 \times \text{presnosť} \times \text{citlivosť}}{\text{presnosť} + \text{citlivosť}}$
F_β štatistika	F_β score	$\frac{(1 + \beta^2) \times \text{presnosť} \times \text{citlivosť}}{\beta^2 \times \text{presnosť} + \text{citlivosť}}$

Zdroj: (Bali, Sarkar, & Lantz, 2016). Pozn.: ^a TP je počet prípadov, ktoré sú pozitívne ($y = 1$) a aj klasifikované ako rizikové ($\hat{y} = 1$); TN je počet prípadov, ktoré sú negatívne ($y = 0$) a aj klasifikované ako nerizikové ($\hat{y} = 0$); FP je počet prípadov, ktoré sú negatívne, ale klasifikované sú ako rizikové; FN je počet prípadov, ktoré sú pozitívne, ale klasifikované sú ako nerizikové. ^b Citlivosť sa dá interpretovať ako správnosť klasifikácie pre skutočne pozitívne prípady ($y = 1$). ^c Špecificita sa dá interpretovať ako správnosť klasifikácie pre skutočne negatívne prípady ($y = 0$).

Vyššie spomínané miery úspešnosti klasifikačného modelu pracujú s binárnym výstupom, ktorý model poskytne až po zvolení hraničnej rizikovosti ρ . Kvalita modelu sa však často posudzuje už pred touto voľbou, pre rôzne hodnoty ρ súčasne. Grafickým nástrojom na porovnanie je tzv. **ROC krivka** (*receiver operating characteristic*), ktorá znázorňuje podiel falošne pozitívnych (*false positive rate*; $FP / (FP + TN)$) a citlivosť, teda podiel skutočne pozitívnych (*true positive rate*), pre rôzne úrovne hraničnej rizikovosti ρ (Obrázok 13). Pre jednoduchosť porovnania modelov sa často používa plocha pod ROC krivkou, **AUC** (*area under curve*), ktorá sumarizuje krivku do jednej numerickej hodnoty. Ideálny model má nulový podiel falošne pozitívnych a stopercentnú citlivosť, čomu zodpovedá bod (0, 1) v grafe a hodnota $AUC = 1$.

5 Úspešnosť modelu

Výstup modelu $p(x)$ sa dá priamo interpretovať ako rizikovosť UoZ s charakteristikami x , známymi v čase jeho zaradenia do evidencie, z hľadiska toho, že sa do jedného roka neuplatní na trhu práce. Pre praktické využitie modelu môže byť potrebný aj výstup v tvare binárnej klasifikácie, kde daného uchádzača o zamestnanie klasifikujeme ako rizikového ($\hat{y} = 1$) alebo nerizikového ($\hat{y} = 0$) nasledovne

$$\hat{y}_i = \begin{cases} 1, & \text{ak } p(x_i) \geq \rho, \\ 0, & \text{ak } p(x_i) < \rho, \end{cases}$$

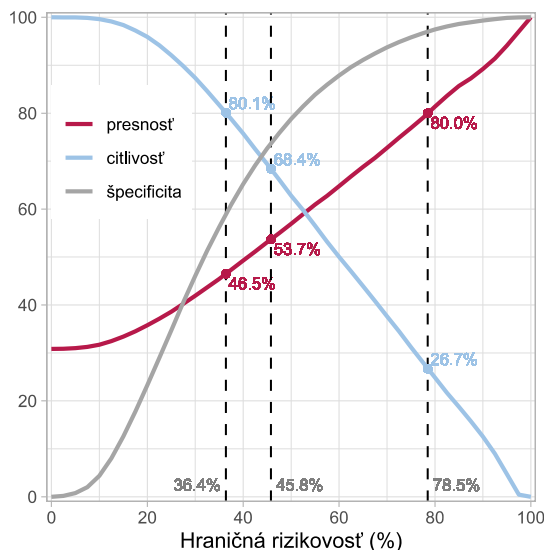
kde **hraničná rizikovosť** $\rho \in [0, 100\%]$, od ktorej je UoZ klasifikovaný ako rizikový, je nastaviteľným parametrom.

Nastavenie hraničnej rizikovosti ρ závisí od viacerých faktorov vrátane vnímania relatívnej dôležitosti citlivosti a presnosti (Box 2), konkrétneho použitia výstupov modelu alebo prípadných kapacitných obmedzení v prípade, že by rizikovým UoZ bola venovaná väčšia pozornosť na úradoch práce. Voľba vysokej hraničnej rizikovosti spôsobí, že len pomerne malá časť UoZ bude označená za rizikových. To povedie k vysokej presnosti modelu (bude málo falošne nerizikových uchádzačov) ale aj k nízkej citlivosti (model zachytí len malú časť skutočne rizikových UoZ). Naopak, nízka hodnota ρ vedie k modelu s nízkou presnosťou ale vysokou citlivosťou (Obrázok 15).

Ak od modelu požadujeme, aby medzi rizikovými UoZ bolo aspoň 80 % takých, ktorí skutočne nenájdu do roka uplatnenie na trhu práce, hraničná pravdepodobnosť bude $\rho = 78,5 \%$ a model podľa *out-of-sample* odhadu na testovacej vzorke zachytí len približne štvrtinu všetkých UoZ s dlhodobým problémom s uplatnením sa (Tabuľka 5). Naopak, ak požadujeme, aby bolo za rizikových označených aspoň 80 % všetkých UoZ, ktorí do roka nenájdu uplatnenie, hraničná pravdepodobnosť bude $\rho = 36,4 \%$ a presnosť bude približne 47 %. **Kompromisná voľba $\rho = 45,8 \%$** , ktorá maximalizuje vyváženú správnosť aj F_1 štatistiku, **vedie k presnosti 53,7 % a citlivosti 68,4 %**.

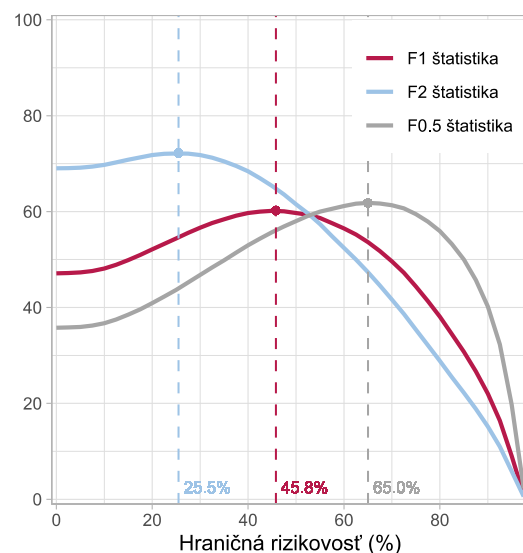
Pre porovnanie, v náhodne zvolenej skupine UoZ sa dá očakávať 30,8 % takých uchádzačov, ktorí si do roka nenájdu uplatnenie (Obrázok 6), kým voľbou pomocou nášho modelu ich bude 53,7 %. Použitím jednoduchého pravidla na identifikáciu rizikových UoZ, ako napríklad nižšie ako stredoškolské vzdelanie alebo prítomnosť aspoň troch znevýhodnení sa dá dosiahnuť presnosť okolo 60 % (Obrázok 7, Obrázok 11), ale takéto klasifikácie odhalia len tretinu, resp. 4,5 % všetkých UoZ, ktorí do roka nenájdu uplatnenie. Naš model z nich odhalí viac ako dve tretiny a aj pri presnosti 60 % by to stále bolo 58 % (Obrázok 15).

Obrázok 15 Presnosť, citlivosť a špecificita v závislosti od hraničnej rizikovosti ρ



Zdroj: vlastné spracovanie

Obrázok 16 F_1 , F_2 a $F_{0,5}$ štatistiky v závislosti od hraničnej rizikovosti ρ



Zdroj: vlastné spracovanie

Tabuľka 5 Voľba hraničnej rizikovosti ρ podľa rôznych mier úspešnosti modelu

Kritérium voľby ρ	ρ	presnosť	citlivosť	špecificita	správnosť	vyvážená správnosť	F_1	F_2	$F_{0,5}$
Presnosť aspoň 80 %	78,5	80,0	26,7	97,0	75,3	61,9	40,0	30,8	57,2
Citlivosť aspoň 80 %	36,4	46,5	80,1	58,9	65,5	69,5	58,9	70,0	50,8
Maximálna správnosť	65,6	69,3	43,2	91,5	76,6	67,3	53,2	46,7	61,8
Maximálna vyvážená správnosť	45,8	53,7	68,4	73,7	72,1	71,1	60,2	64,9	56,1
Maximálna F_1	45,8	53,7	68,4	73,7	72,1	71,1	60,2	64,9	56,1
Maximálna F_2	25,5	38,9	91,7	35,9	53,1	63,8	54,7	72,2	44,0
Maximálna $F_{0,5}$	65,0	68,8	43,9	91,1	76,6	67,5	53,6	47,4	61,8

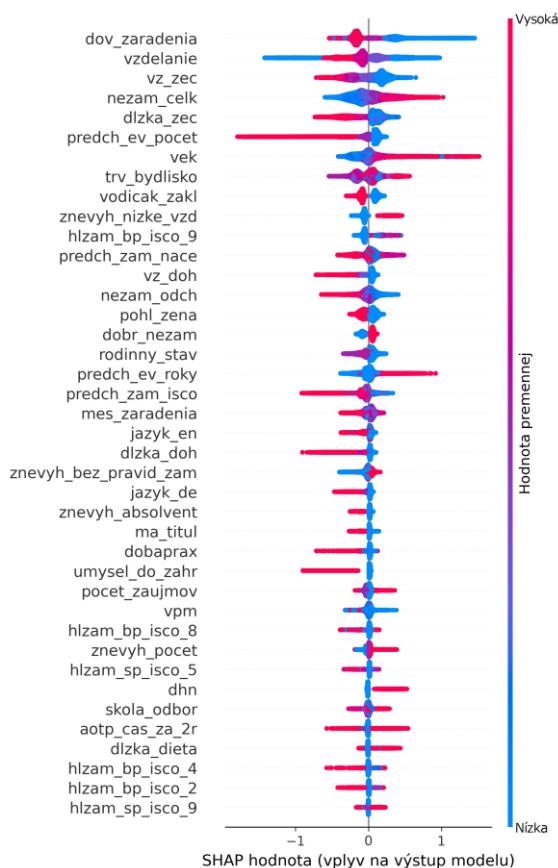
Zdroj: vlastné spracovanie. Pozn.: Všetky hodnoty sú uvedené v percentách.

6 Prediktory rizikovosti

Najvýznamnejšie premenné pri posudzovaní rizikovosti uchádzača o zamestnanie z hľadiska neuplatnenia sa na trhu práce do jedného roka od zaradenia do evidencie sú dôvod zaradenia do evidencie, úroveň dosiahnutého vzdelania a priemerná výška príjmu zo zamestnaneckých poisťných vzťahov v období dvoch rokov pred zaradením do evidencie (Obrázok 17).

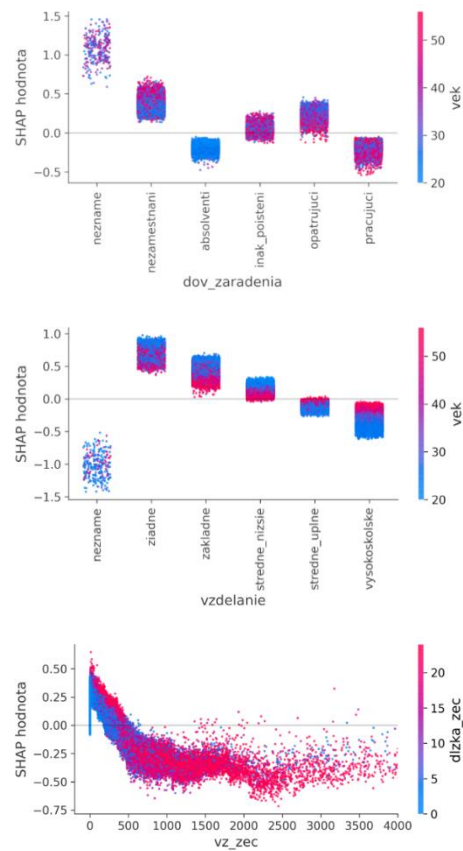
Na posúdenie dôležitosti premenných aj na interpretáciu vplyvu hodnôt premenných na výstup modelu používame prístup SHAP (SHapley Additive exPlanations), ktorý využíva Shapley hodnoty a princípy z teórie kooperatívnych hier. SHAP hodnota pre i -teho UoZ a j -tu vysvetľujúcu premennú kvantifikuje, ako konkrétna hodnota j -tej premennej prispieva k odchýlke rizikovosti i -teho UoZ od priemernej rizikovosti všetkých UoZ v tréningovej vzorke. Finálnu mieru rizikovosti i -teho UoZ tvorí súčet SHAP hodnôt pre všetky vysvetľujúce premenné a priemernej, resp. základnej rizikovosti všetkých UoZ v tréningovej vzorke.¹⁷ Dôležitosť j -tej vysvetľujúcej premennej určuje priemerná absolútna hodnota SHAP hodnôt tejto premennej pre jednotlivých UoZ. Čím väčšie odchýlky hodnota danej premennej spôsobuje na celkovej miere rizikovosti UoZ, tým je táto premenná dôležitejšia pre model.¹⁸

Obrázok 17 Najsilnejšie prediktory rizikovosti sú dôvod zaradenia do evidencie, dosiahnuté vzdelanie a výška predchádzajúceho príjmu zo zamestnania



Zdroj: vlastné spracovanie pomocou shap.summary_plot.
Pozn.: Farba bodu zodpovedá hodnote premennej pre konkrétneho UoZ, horizontálna poloha vplyvu na jeho rizikovosť a vertikálny rozptyl bodov koncentrácií podobných hodnôt.

Obrázok 18 SHAP hodnoty pre tri najsilnejšie prediktory¹⁹



Zdroj: vlastné spracovanie pomocou shap.dependence_plot.
Pozn.: Každý bod zodpovedá konkrétnemu UoZ. Horizontálna poloha zodpovedá hodnote zvolenej vysvetľujúcej premennej, vertikálna poloha vplyvu tejto hodnoty na rizikovosť. Farba zohľadňuje hodnotu inej zvolenej premennej a pomáha ilustrovať interakcie medzi premennými.

Pri **dôvode zaradenia do evidencie** pozorujeme konzistentne negatívny vplyv na rizikovosť (zníženie rizikovosť) v skupinách pracujúcich a absolventov škôl (Obrázok 18, horný panel). Naopak, medzi osobami, ktoré pred zaradením nepracovali alebo sa starali o dieťa alebo blízku osobu je rizikovosť zvýšená. Pri nepracujúcich vyšší vek skôr zvyšuje rizikovosť, kým pri opatrujúcich znižuje. Neuvedenie dôvodu zaradenia má najväčší a pozitívny vplyv na rizikovosť (zvyšuje rizikovosť).

¹⁷ SHAP hodnoty nevyjadrujú rizikovosť v jednotkách pravdepodobnosti p ale v logaritmickej šanciach ($\log\text{-odds}$) $\ln \frac{p}{1-p}$.

¹⁸ Bližšie informácie o SHAP hodnotách môže čitateľ nájsť v literatúre (Lundberg & Lee, 2017; Molnar, 2019).

¹⁹ SHAP hodnoty pre ostatné použité premenné sú zobrazené v prílohe B.

Vyššie **vzdelanie** znižuje riziko neuplatnenia sa na trhu práce do jedného roka, pričom vplyv dosiahnutého vzdelania je vyšší pri mladších UoZ ako pri starších (Obrázok 18, stredný panel).²⁰ Neuvedenie úrovne vzdelania prekvapivo výrazne znižuje rizikovosť, čo môže súvisieť s nízkou motiváciou vyplňať údaje v žiadosti o zaradenie do evidencie pri uchádzačoch, ktorí očakávajú, že sa bez problémov zamestnajú.²¹

Pri **úrovni predchádzajúcich príjmov zo zamestnania** pozorujeme konzistentný negatívny vplyv na rizikovosť pri mesačných príjmoch zhruba od úrovne 500 eur a to bez ohľadu na dĺžku poberania týchto príjmov (Obrázok 18, spodný panel). Naopak, rizikovosť narastá hlavne osobám s nízkym predchádzajúcim zamestnaneckým príjmom a tento vplyv je silnejší, ak takéto príjem poberali dlhší čas.

Vplyv **miery nezamestnanosti** v okrese v čase registrácie má lineárny a rastúci charakter a vplyv je silnejší pre ženy (príloha B). Prítomnosť **zamestnaneckých pracovných vzťahov** v menej ako polovici dvojročného obdobia pred zaradením do evidencie zvyšuje rizikovosť, kým prítomnosť zamestnaneckých vzťahov aspoň počas roku a pol rizikovosť znižuje. Prekvapivý je vplyv **počtu predchádzajúcich evidencií** na úrade práce. Nulový počet rizikovosť mierne zvyšuje a rastúci počet vedie k silnejšiemu zápornému vplyvu.²² Vplyv obdobia zamestnaneckých vzťahov aj počtu predchádzajúcich evidencií je silnejší pre starších UoZ.

Z hľadiska **veku** uchádzača je rizikový vek pod 18 rokov, čo môže súvisieť s nízkym, prípadne neukončeným vzdelaním a chýbajúcimi pracovnými skúsenosťami takýchto UoZ. Inak nízky vek rizikovosť skôr znižuje a okolo 35 rokov sa vplyv prekloní na zvyšovanie. Okres **trvalého bydliska** sme pre každý kraj rozdelili na okresy krajského mesta a skupinu ostatných okresov v kraji. Pri všetkých krajských mestách s výnimkou Žiliny vyšiel vplyv na rizikovosť kladný, pričom najvýraznejšie zvyšuje rizikovosť trvalý pobyt v Košiciach. Naopak, negatívny vplyv na rizikovosť má len trvalé bydlisko v Trenčianskom kraji okrem Trenčína, celom Žilinskom kraji a v Prešovskom kraji okrem Prešova.²³

Dobrymi prediktormi rizikivosti sú aj **základný vodičský preukaz, znevýhodnenie nízkym vzdelaním** (obzvlášť pre mladších UoZ) alebo **záujem o zamestnanie ako pomocný alebo nekvalifikovaný pracovník bez praxe**.²⁴ Vplyv pohlavia na rizikovosť je na východe nižší ako na západe, vplyv minulého obdobia dobrovoľnej nezamestnanosti je silnejší pre mladších uchádzačov a vplyv rodinného stavu je výraznejší pre mužov. **Informácie o počítačových zručnostiach vyplnené pri registrácii uchádzačov sa prekvapivo neukázali pri modelovaní ako dôležité** a do konečného modelu sa ani nedostali.

7 Modelové príklady určenia rizikivosti

Obrázok 19 ilustruje pre dvoch náhodne zvolených uchádzačov priradenie rizikivosti, že osoba nenájde uplatnenie na trhu práce do jedného roka od zaradenia do evidencie UoZ. Ľavý panel ukazuje uchádzačku UoZ_1 , ktorá má 20 rokov a je absolventkou s úplným stredoškolským vzdelaním v Prešovskom kraji. Na pravom paneli je zobrazený uchádzač UoZ_2 z Nitrianskeho kraja, ktorý má 21 rokov, základné vzdelanie a pred zaradením do evidencie nepracoval. Prvá uchádzačka skutočne našla uplatnenie do jedného roka a model jej

²⁰ Tieto výsledky sú v súlade s prvotnými pozorovaniami dát (Obrázok 7).

²¹ Táto hypotéza sa nepotvrdzuje pri odbore vzdelania, kde neuvedenie hodnoty nemá jednoznačný vplyv na rizikovosť. Dôkladné vyhodnotenie vzťahu medzi mierou vyplnenia registračného dotazníka UoZ a pravdepodobnosti jeho včasného uplatnenia sa na trhu práce by bolo vhodné analyzovať v samostatnej štúdií venovanej tejto otázke. Táto výskumná otázka je však nad rámec aktuálneho výstupu.

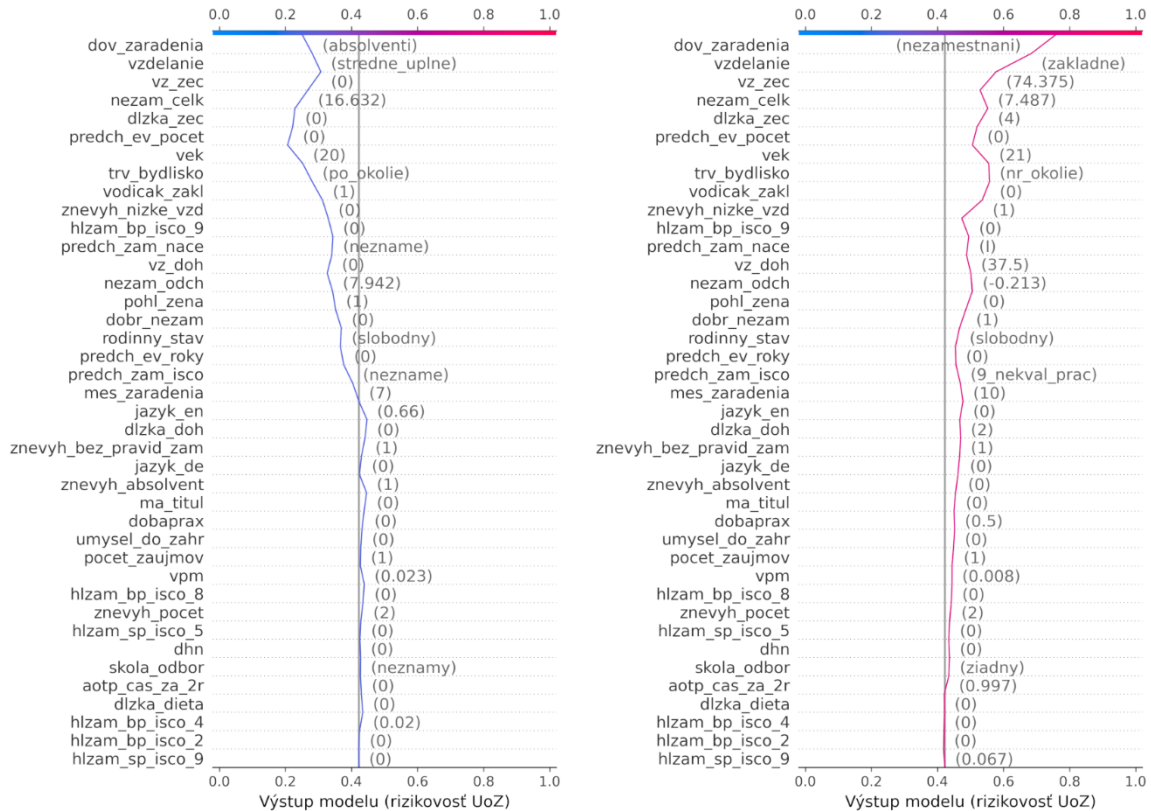
²² Znižovanie rizikivosti neuplatnenia sa na trhu práce pri vysokom počte predchádzajúcich evidencií v posledných dvoch rokoch zrejme súvisí s tým, že takáto osoba musela byť aj mnohokrát vyradená z evidencie, a preto sa „nekvalifikovala“ ako riziková.

²³ Nízka rizikovosť Žiliny a vysoká rizikovosť Košíc bola očakávaná aj na základe pozorovaní na Obrázku 11. Naopak, nízka rizikovosť okresov v Prešovskom kraji z týchto pozorovaní jasná nebola.

²⁴ Premenná *hlzam_bp_isco_9* má pre konkrétneho UoZ hodnotu 0, ak neuviedol záujem o pozíciu v hlavnej triede 9 podľa SK-ISCO 08, alebo uviedol okrem záujmu aj prax na takejto pozícii. V prípade, že UoZ uviedol záujem, ale neuviedol prax, hodnota tejto premennej predstavuje podiel evidovaných voľných pracovných miest v okrese a čase registrácie na jedného UoZ so záujmom o prácu na takejto pozícii.

správne priradil nízku rizikovosť 25,2 %. Druhý UoZ uplatnenie nenašiel a model mu tiež správne priradil pomerne vysokú rizikovosť 75,6 %.

Obrázok 19 Správne klasifikovani uchádzači s rizikovosťami 25,2 % (UoZ₁, ľavý panel) a 75,6 % (UoZ₂, pravý panel)

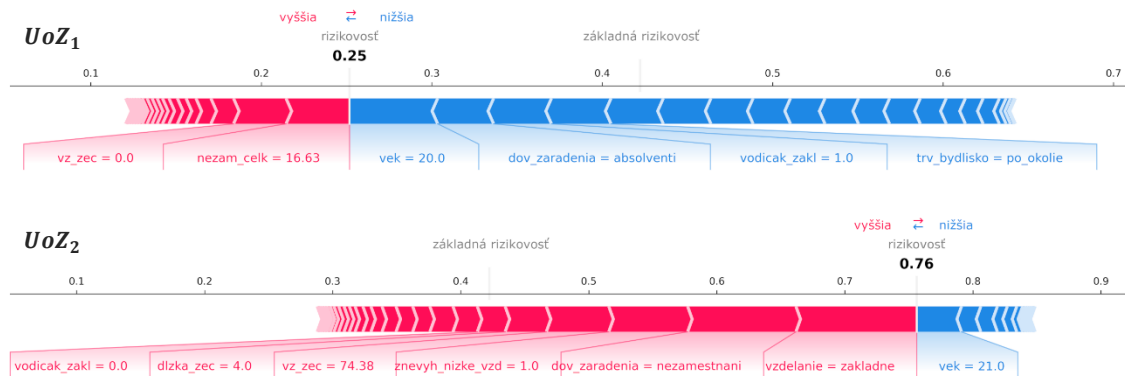


Zdroj: vlastné spracovanie pomocou shap.decision_plot. Pozn.: Premenné sú zoradené podľa celkovej dôležitosti pre model. V zátvorkách sú zobrazené konkrétne hodnoty premenných pre zvolených UoZ.

Vplyv jednotlivých charakteristík uchádzača na jeho výslednú rizikovosť nie je pre všetkých rovnaký. Pre UoZ₁ SHAP hodnoty ukazujú, že pri znižovaní jej rizikovosti je najdôležitejšie, že má 20 rokov, je absolventka, má vodičský preukaz a trvalé bydlisko v Prešovskom kraji (Obrázok 20). Naopak, jej rizikovosť najviac zvyšujú miera nezamestnanosti 16,63 % v okrese jej trvalého bydliska a nulový predchádzajúci príjem zo zamestnaneckej činnosti. Rizikovosť druhého uchádzača najviac znižuje jeho vek 21 rokov. Faktorov, ktoré vedú k celkovej vysokej rizikovosti je niekoľko – má len základné vzdelanie, pred zaradením do evidencie nepracoval, je znevýhodnený nízkym vzdelaním, v posledných dvoch rokoch mal zamestnanecký pomer len 4 mesiace s priemernou mesačnou mzdou 74 eur a nemá vodičský preukaz.

Obrázok 21 v Prílohe C ukazuje dva prípady, v ktorých model uchádzačov klasifikuje nesprávne. Ľavý panel zobrazuje 26-ročnú ženu z Nitry s vysokoškolským vzdelaním, ktorá 20 mesiacov z posledných dvoch rokov pred zaradením do evidencie pracovala s priemernou mesačnou mzdou 1011 eur. Model jej na základe jej charakteristík priradil nízku rizikovosť 11,7 %, na čom sa najviac podieľala jej predchádzajúca mzda, vysokoškolské vzdelanie, miera nezamestnanosti mierne pod 4 % v jej okrese v čase registrácie a fakt, že pred zaradením pracovala (Obrázok 22). V skutočnosti sa však táto uchádzačka do roka od zaradenia do evidencie UoZ na trhu práce neuplatnila. Na pravom paneli Obrázku 20 je zobrazený 30-ročný muž z Banskobystrického kraja, ktorý má základné vzdelanie a pred zaradením do evidencie nepracoval. Model posúdil jeho rizikovosť na takmer 81 %, no v skutočnosti tento UoZ našiel uplatnenie skôr ako za 12 mesiacov.

Obrázok 20 Najdôležitejšie charakteristiky pre uchádzačov UoZ₁ a UoZ₂ z obrázka 18



Zdroj: vlastné spracovanie pomocou shap.force_plot. Pozn.: Červenou sú znázornené charakteristiky, ktoré najviac zvyšujú rizikovosť daného uchádzača, modrou charakteristiky, ktoré rizikovosť najviac znižujú.

8 Záver

Dlhodobá nezamestnanosť je jednou z pretrvávajúcich príčin chudoby a má výrazné dôsledky na spoločnosť, ekonomiky krajín EÚ a nespochybniteľne aj na samotného jednotlivca. (Scoppetta & Buckenleib, 2018). Medzinárodne najlepšie overenú prax, ako bojovať s jednou z najzávažnejších výziev politiky trhu práce nielen na Slovensku, tvorí kvantitatívna profilácia UoZ, ktorá je zameraná na predchádzanie dlhodobej nezamestnanosti.

Exploratívna analýza odhalila skupiny populácie, v ktorých je problém s nájdením pracovného uplatnenia častejší. Medzi takéto skupiny patria UoZ s nízkym vzdelaním, s vyšším počtom znevýhodnení, bez vodičského oprávnenia alebo bez počítačových znalostí. Stále pretrvávajúce regionálne rozdiely sú ľahko pozorovateľné pri vyššej koncentrácii osôb s dlhodobým problémom s pracovným uplatnením v Banskobystrickom, Prešovskom a Košickom kraji. Prekvapivým zistením je, že medzi absolventami škôl, ktorí sú považovaní za znevýhodnených UoZ, je problém s uplatnením sa zriedkavejší ako vo zvyšku populácie.

Výsledky modelu ukazujú, že rizikovosť uchádzačov najviac zvyšuje absencia pracovného pomeru bezprostredne pred zaradením do evidencie, nízke vzdelanie, trvalé bydlisko v okrese s vysokou nezamestnanosťou, dlhotrvajúci pracovný pomer s nízkou mzdou alebo krátko trvajúci zamestnanecký pomer v posledných dvoch rokoch pred zaradením do evidencie. Naopak medzi faktory spôsobujúce pokles rizikovosti uchádzačov patrilo najvyššie dosiahnuté vzdelanie aspoň na úrovni úplného stredoškolského vzdelania, zamestnanecký pomer s výškou mzdy aspoň 500 € mesačne, či existencia pracovného pomeru alebo absolvovanie školy bezprostredne pred zaradením do evidencie UoZ. Informácie o počítačových znalostiach vyplnené pri registrácii uchádzačov naopak nenapomáhajú k lepšiemu odhadnutiu ich rizikovosti.

Významnou výhodou profilácie UoZ pomocou strojového učenia je flexibilita pri identifikácii rizikových UoZ. S využitím dostupných dát a navrhovaného nastavenia modelu dokážeme pomocou metódy strojového učenia, LightGBM, identifikovať viac ako dve tretiny uchádzačov, ktorí majú problém s uplatnením sa na trhu práce, pričom viac ako polovica UoZ označených ako rizikových bude označená správne priamo pri zaradení do evidencie UoZ. Model v prípade potreby umožňuje operatívne pozmeniť hodnotu hraničnej rizikovosti, čo umožňuje koncentrovať sa iba na identifikáciu veľmi rizikových prípadov zvýšením presnosti modelu alebo na zachytenie väčšieho počtu rizikových UoZ zvýšením citlivosti.

Jednou z hlavných predností využívania systému na predikciu rizikovosti upadnutia do dlhodobej nezamestnanosti je dátami podložená **včasná identifikácia problematických UoZ** už pri zaradení do evidencie. Nástroje na profiláciu pomáhajú **zvyšovať nákladovú efektívnosť** tým, že umožňujú vyššie zameranie sa na rizikových UoZ a zníženie nákladov na UoZ, ktorí nebudú mať problém nájsť si uplatnenie na trhu práce. V neposlednom rade prináša **zvýšenie úrovne poskytovaných služieb**, keďže dáva možnosť

prispôbiť sa individuálnym potrebám UoZ (Desiere, Langenbucher, & Struyven, 2019). Posúdenie rizikivosti rovnakým modelom na všetkých úradoch práce na území Slovenskej republiky by mohlo nahradiť subjektívne rozhodnutia pracovníkov objektívnym úsudkom založeným na získaných dátach.

Vybraný sofistikovaný model LightGBM dosahuje vyššiu úspešnosť ako jednoduchšie modely typu logistickej regresie častejšie využívané v iných krajinách. Pomerne zložitejšiu interpretovateľnosť a nižšiu transparentnosť modelu oproti jednoduchším modelom **preto zvyšujeme použitím SHAP hodnôt**, ktoré okrem identifikácie celkovo najdôležitejších charakteristík pre všetkých UoZ umožňujú aj **pre konkrétnu osobu posúdiť, aké sú jej silné a slabé stránky z hľadiska rizikivosti**. Takýto druh informácií môže byť obzvlášť užitočný pri účinnej reintegrácii uchádzača na pracovnom trhu.

Medzinárodná vedecká literatúra poukazuje na **prínos** tzv. **mäkkých dát** v procese profilácie UoZ (Troya, a iní, 2018). Tieto poskytujú informáciu o osobnostných charakteristikách UoZ, ako je motivácia, názor na nezamestnanosť alebo aj uvedomovanie si svojich vlastných zručností, ktoré ale v použitých dátach neboli zastúpené. V aktuálnej žiadosti o zaradenie do evidencie platnej od júla 2020 sa tento trend však už stihol premietnuť. UoZ označí, ktoré dôvody vníma ako prekážku pri uplatnení na trhu práce, čo môže výstižne napovedať o vnímaní jeho vlastných zručností alebo názoru. Na druhej strane, zo žiadosti bola vypustená celá časť zameriavajúca sa na záujmy a úmysly UoZ. Za následok to má vypadnutie dvoch premenných z nášho modelu, ktorými sú počet uvedených záujmov a úmysel odísť do zahraničia.

Úspešne implementovaná profilácia UoZ by tiež ponúkla nástroj na jednoduché hodnotenie účinnosti konkrétnych politik trhu práce na zlepšenie uplatniteľnosti UoZ. Zohľadnením absolvovania daného opatrenia v modeli by bolo možné posúdiť, či táto politika má vplyv na rizikivosť a prípadne ho aj kvantifikovať. Takéto posúdenie politik by nenahrádzalo komplexnejšie analýzy zohľadňujúce aj náklady a prínosy, ale umožnilo by jednoducho zhodnotiť, či daná politika vedie k rýchlejšiemu uplatneniu UoZ na trhu práce.

Na základe analýzy odporúčame

- **zaviesť systematickú včasnú profiláciu uchádzačov o zamestnanie** na základe štatistickej analýzy dát, ktorá pomôže identifikovať rizikových UoZ;
- **zabezpečiť kontinuitu sledovania významných charakteristík UoZ** – vítame začatie zberu aj mäkkých dát o UoZ, ale z hľadiska úspešnej profilácie je dôležitá aj kontinuálna dostupnosť všetkých údajov, ktoré sa ukazujú ako významné pri posudzovaní rizikivosti uchádzačov;
- **rozpracovať možnosti hodnotenia rizikivosti pre špecifické demografické skupiny**, akými sú napríklad viacnásobne znevýhodnení UoZ, uchádzači so zdravotným postihnutím, UoZ v pred dôchodkovom veku, absolventi alebo UoZ, ktorí si neudržia zamestnanie a opakovane sa objavujú v evidencii.

Literatúra

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . Google Brain. (2016). Tensorflow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, (s. 265-283). Savannah, GA, USA.
- Bali, R., Sarkar, D., & Lantz, B. (2016). *R: Unleash Machine Learning Techniques*. Birmingham: Packt Publishing.
- Desiere, S., Langenbucher, K., & Struyven, L. (2019). *Statistical profiling in public employment services: An international comparison*. Paríž: OECD Social, Employment and Migration Working Papers, No. 224, OECD Publishing.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2. vyd.). New York: Springer.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *22nd SIGKDD Conference on Knowledge Discovery and Data Mining*, (s. 785–794). San Francisco, CA, USA.
- Chen, T., He, T., Benesty, M., & Tang, Y. (2020). *Understand your dataset with Xgboost*. Dostupné na Internete: <https://cran.r-project.org/web/packages/xgboost/vignettes/discoverYourData.html>
- Ke, G., Meng, Q., Finely, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*.
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems, 30*, 4765-4774.
- Molnar, C. (2019). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Dostupné na Internete: <https://christophm.github.io/interpretable-ml-book/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.
- R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Dostupné na Internete: <https://www.R-project.org/>
- Scoppetta, A., & Buckenleib, A. (2018). *Tackling Long-Term Unemployment through Risk Profiling and Outreach*. Brusel: ESF Transnational Platform.
- Soukup, T. (2011). Profiling: Predicting Long-Term Unemployment at the Individual Level. *Central European Journal of Public Policy, 118-142*.
- Troya, Í. M., Chen, R., Moraes, L. O., Bajaj, P., Kupersmith, J., Ghani, R., . . . Zejnilovic, L. (2018). Predicting, explaining, and understanding risk of long-term unemployment. *32nd Conference on Neural Information Processing Systems*. Montreal.
- Van Rossum, G., & Drake Jr, F. L. (1995). *Python tutorial*. Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica.
- Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open, 4(43)*, 1686.

Prílohy

A Vysvetľujúce premenné

Tabuľka 6 Vysvetľujúce premenné použité vo výslednom modeli

#	Premenná	Popis	Počet kategórií / priemer ^a	Najpočetnejšia kategória / podiel núl ^b
Údaje z evidencie UoZ (ÚPSVaR)				
1	vek	vek zaokrúhlený na celé roky	priem. 34	0%
2	pohl_zena	pohlavie: muž = 0 / žena = 1	2	0 (53,25 %)
3	trv_bydlisko	trvalé bydlisko: okres krajského mesta / iný okres v kraji	16	po_okolie (15,70 %)
4	vzdelanie	vzdelanie: žiadne / základné / stredné nižšie / stredné úplné / vysokoškolské / neznáme	6	stredne_uplne (37,53 %)
5	skola_odbor	odbor vzdelania: kultúra / neznámy / poľnohospodárstvo / prírodné vedy / spoločenské vedy 1 ...	11	neznamy (28,45 %)
6	ma_titul	akademický titul pred alebo za menom: nie = 0 / áno = 1	2	0 (81,43 %)
7	rodinny_stav	rodinný stav: rozvedený / slobodný / vdovec / záväzok	4	slobodny (57,44 %)
8	vodicak_zakl	základný vodičský preukaz (skupiny A, A1, A2, AM, B a B1): nie = 0 / áno = 1	2	1 (51,71 %)
9	jazyk_en	úroveň znalosti anglického jazyka: žiadny = 0 / A = 0,33 / B = 0,66 / C = 1	4	0 (59,22 %)
10	jazyk_de	úroveň znalosti nemeckého jazyka: žiadny = 0 / A = 0,33 / B = 0,66 / C = 1	4	0 (73,50 %)
11	mes_zaradenia	mesiac, v ktorom bol uchádzač zaradený do evidencie UoZ	12	9 (11,57 %)
12	dov_zaradenia	dôvod zaradenia do evidencie: absolvent / inak poistený / nezamestnaný / neznámy / opatrujúci / pracujúci	6	pracujuci (51,82 %)
13	predch_zam_nace	odvetvie predchádzajúceho zamestnania podľa SK-NACE	14	nezname (36,49 %)
14	prech_zam_isco	kód hlavnej triedy predchádzajúceho zamestnania podľa SK ISCO-08	10	nezname (22,18 %)
15	predch_ev_pocet	počet predchádzajúcich zaradení do evidencie za 2 roky pred aktuálnym zaradením	priem. 1,36	60,38 %
16	predch_ev_roky	počet rokov strávených v evidencii UoZ za 2 roky pred aktuálnym zaradením	priem. 0,23	60,39 %
17	dobr_nezam	bol UoZ v minulosti dobrovoľne nezamestnaný: nie = 0 / áno = 1	2	1 (60,56 %)
18	umysel_do_zahr	má UoZ úmysel odísť do zahraničia? nie = 0 / áno = 1	2	0 (97,06 %)
19	pocet_zaujmov	počet uvedených záujmov: 1 / 2 / 3 / 4 a viac = 4	4	2 (53,79 %)
20	znevych_absolvent	znevýhodnenie - absolvent: nie = 0 / áno = 1	2	0 (85,20 %)
21	znevych_bez_pravid_zam	znevýhodnenie - 12 kal. mes. nemal pravidelne plat. zam.: nie = 0 / áno = 1	2	1 (58,23 %)
22	znevych_nizke_vzd	znevýhodnenie - nízke vzdelanie : nie = 0 / áno = 1	2	0 (84,07 %)
23	znevych_pocet	počet znevýhodnení UoZ: 0 / 1 / 2 / 3 a viac = 3	4	1 (37,75 %)
24	dobaprax	počet rokov praxe v súčte za uvedené zamestnania	priem. 9,69	77,34 %
25	hlzam_sp_isco_5	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 5 podľa SK ISCO-08 v prípade záujmu a praxe	priem. 0,105	82,05 %
26	hlzam_sp_isco_9	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 9 podľa SK ISCO-08 v prípade záujmu a praxe	priem. 0,086	77,67 %
27	hlzam_bp_isco_2	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 2 podľa SK ISCO-08 v prípade záujmu a bez praxe	priem. 0,085	87,55 %
28	hlzam_bp_isco_4	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 4 podľa SK ISCO-08 v prípade záujmu a bez praxe	priem. 0,073	83,77 %
29	hlzam_bp_isco_8	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 8 podľa SK ISCO-08 v prípade záujmu a bez praxe	priem. 0,436	89,40 %
30	hlzam_bp_isco_9	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 9 podľa SK ISCO-08 v prípade záujmu a bez praxe	priem. 0,081	77,51 %
Iné údaje z ÚPSVaR				
31	dhn	bol UoZ poberateľom dávky v hmotnej núdzi? nie = 0 / áno = 1	2	0 (96,60 %)
32	aotp_cas_za_2r	počet rokov strávených na AOTP za 2 roky pred aktuálnym zaradením	priem. 0,684	89,73 %
33	nezam_celk	miera nezamestnanosti z celkového počtu UoZ v okrese 2 mesiace pred aktuálnym zaradením	priem. 11,55	0 %
34	nezam_odch	odchýlka miery nezamestnanosti v okrese od celoslovenskej miery nezamestnanosti 2 mesiace pred aktuálnym zaradením	priem. 1,18	0 %
35	vpm	podiel voľných pracovných miest v okrese na všetkých voľných pracovných miestach v SR 2 mesiace pred aktuálnym zaradením	priem. 0,024	0 %

Údaje zo Sociálnej poisťovne				
36	vz_zec	mesačný priemerný vymeriavací základ zo zamestnaneckých poistných vzťahov za 2 roky pred aktuálnym zaradením	priem. 640 €	38,59 %
37	vz_doh	mesačný priemerný vymeriavací základ z dohodárskych poistných vzťahov za 2 roky pred aktuálnym zaradením	priem. 118 €	68,54 %
38	dlzka_zec	počet mesiacov so zamestnaneckým poistným vzťahom za 2 roky pred aktuálnym zaradením	priem. 14,34	38,59 %
39	dlzka_doh	počet mesiacov s dohodárskym poistným vzťahom za 2 roky pred aktuálnym zaradením	priem. 5,36	68,54 %
40	dlzka_dieta	počet mesiacov strávených na materskej dovolenke alebo starostlivosťou o dieťa za 2 roky pred aktuálnym zaradením	priem. 17,86	93,62 %

Pozn: ^a Pre kategorické premenné uvádzame počet kategórií, pre spojité priemer z nenulových hodnôt. ^b Pre kategorické premenné uvádzame zastúpenie najpočetnejšej kategórie, pre spojité podiel nulových hodnôt.

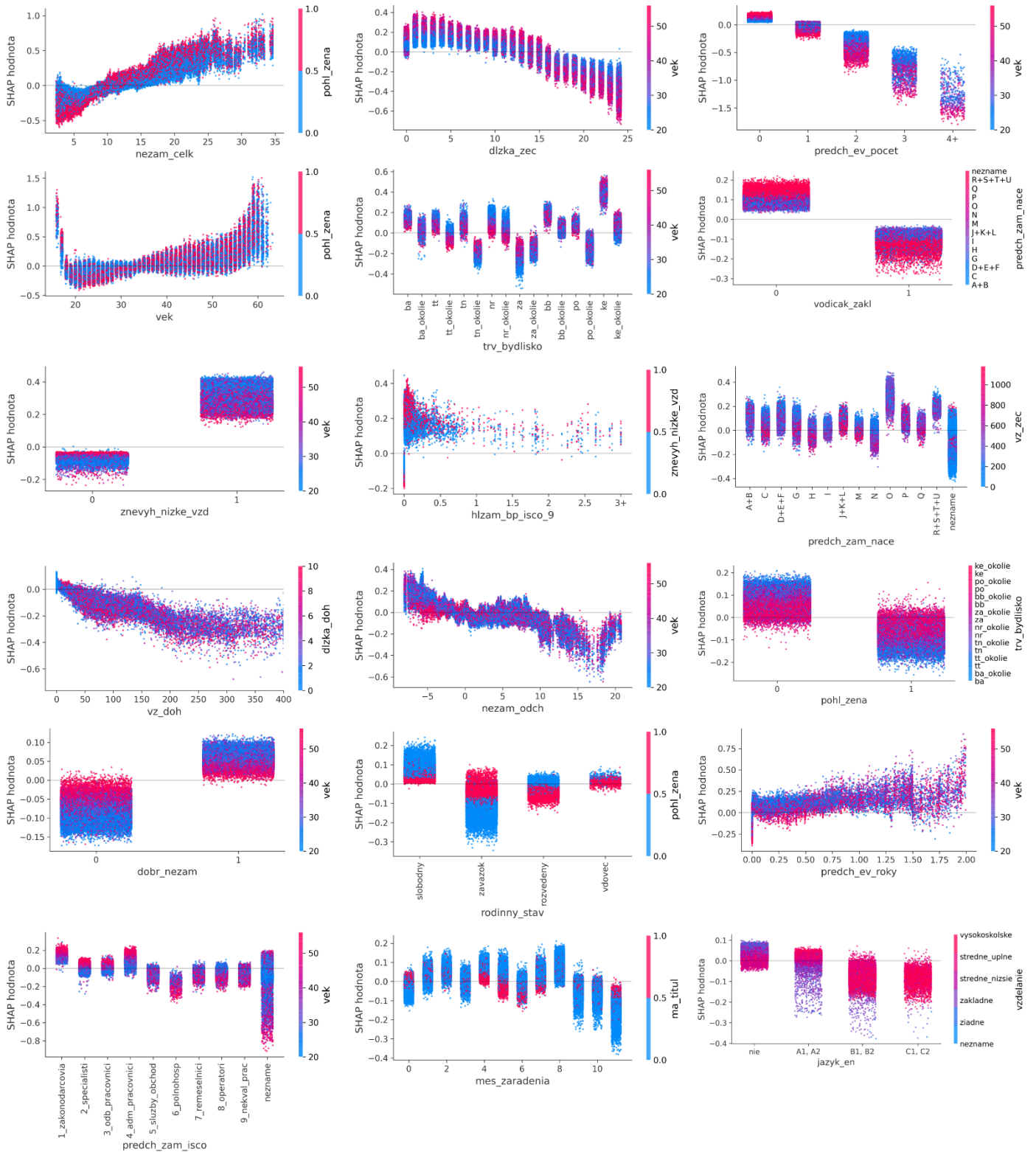
Tabuľka 7 Vysvetľujúce premenné, ktoré neboli vybraté do výsledného modelu

#	Premenná	Popis	Počet kategórií / priemer ^a	Najpočetnejšia kategória / podiel núl ^b
Údaje z evidencie UoZ (ÚPSVaR)				
1	prech_bydlisko	prechodné bydlisko: okres s nízkou nezamestnanosťou / iný okres / neurčené	3	neurcene (95,27 %)
2	pocet_deti	počet detí: 0 / 1 / 2 / 3 a viac = 3	4	0 (90,37 %)
3	vodicak_prof	profesionálny vodičský preukaz (skupiny BE, C, CE, C1, C1E, D, DE, D1, D1E, T): nie = 0 / áno = 1	2	0 (92,38 %)
4	jazyk_sk	znalosť slovenského jazyka: nie = 0 / áno = 1	2	0 (97,26 %)
5	jazyk_hu	znalosť maďarského jazyka: nie = 0 / áno = 1	2	0 (88,19 %)
6	jazyk_pl	znalosť poľského jazyka: nie = 0 / áno = 1	2	0 (98,83 %)
7	jazyk_iny_svet	znalosť iného svetového jazyka (francúzsky, španielsky, taliansky ...): nie = 0 / áno = 1	2	0 (95,65 %)
8	jazyk_iny	znalosť iného jazyky: nie = 0 / áno = 1	2	0 (92,94 %)
9	jazyk_ru	úroveň znalosti ruského jazyka: žiadny = 0 / A = 0,33 / B = 0,66 / C = 1	4	0 (91,79 %)
10	pc_internet	znalosť ovládania pc / internet: nie = 0 / áno = 1	2	0 (92,89 %)
11	msoff_zakl_pokroc	Microsoft Office úroveň znalosti základná alebo pokročilá: nie = 0 / áno = 1	2	0 (52,62 %)
12	msoff_vysoka	Microsoft Office úroveň znalosti vysoká: nie = 0 / áno = 1	2	0 (90,59 %)
13	pc_ine	iné počítačové znalosti: nie = 0 / áno = 1	2	0 (92,31 %)
14	narodnost	národnosť: slovenská / maďarská / iná	3	sk (90,52 %)
15	st_prisl	štátna príslušnosť: sr / EÚ / tretie krajiny	2	sk (99,44 %)
16	Szco	bol UoZ v minulosti SZČO: nie = 0 / áno = 1	2	0 (99,82 %)
17	prislub_zam	má UoZ prísľub zamestnania: nie = 0 / áno = 1	2	0 (99,85 %)
18	zaujem_doch_zam	záujem dochádzať do zamestnania: do 50 km / nad 50 km / nie	3	nad 50 km (80,77 %)
19	zaujem_zam_zahr	záujem pracovať aj mimo územia SR: nie = 0 / áno = 1	2	0 (78,92 %)
20	zaujem_szco	záujem vykonávať SZČ: nie = 0 / áno = 1	2	0 (94,36 %)
21	zaujem_skr_uvazok	záujem o iný ako plný úväzok: nie = 0 / áno = 1	2	0 (97,08 %)
22	zaujem_vzdel	záujem o vzdelávanie: nie = 0 / áno = 1	2	0 (73,97 %)
23	zaujem_absprax	záujem o absolventskú prax: nie = 0 / áno = 1	2	0 (95,78 %)
24	zaujem_sust_prip	záujem o sústavnú prípravu na povolanie: nie = 0 / áno = 1	2	0 (97,04 %)
25	znevyh_nad_50_rokov	znevýhodnenie - nad 50 rokov: nie = 0 / áno = 1	2	0 (85,84 %)
26	znevyh_zdravotne	znevýhodnenie - zdravotné: nie = 0 / áno = 1	2	0 (97,98 %)
27	znevyh_ine	znevýhodnenie - iné: nie = 0 / áno = 1	2	0 (99,77 %)
28	hlzam_sp_isco_1	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 1 podľa SK ISCO-08 v prípade záujmu a praxe	priem. 0,071	97,80 %
29	hlzam_sp_isco_2	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 2 podľa SK ISCO-08 v prípade záujmu a praxe	priem. 0,084	93,59 %
30	hlzam_sp_isco_3	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 3 podľa SK ISCO-08 v prípade záujmu a praxe	priem. 0,068	91,81 %
31	hlzam_sp_isco_4	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 4 podľa SK ISCO-08 v prípade záujmu a praxe	priem. 0,074	90,44 %

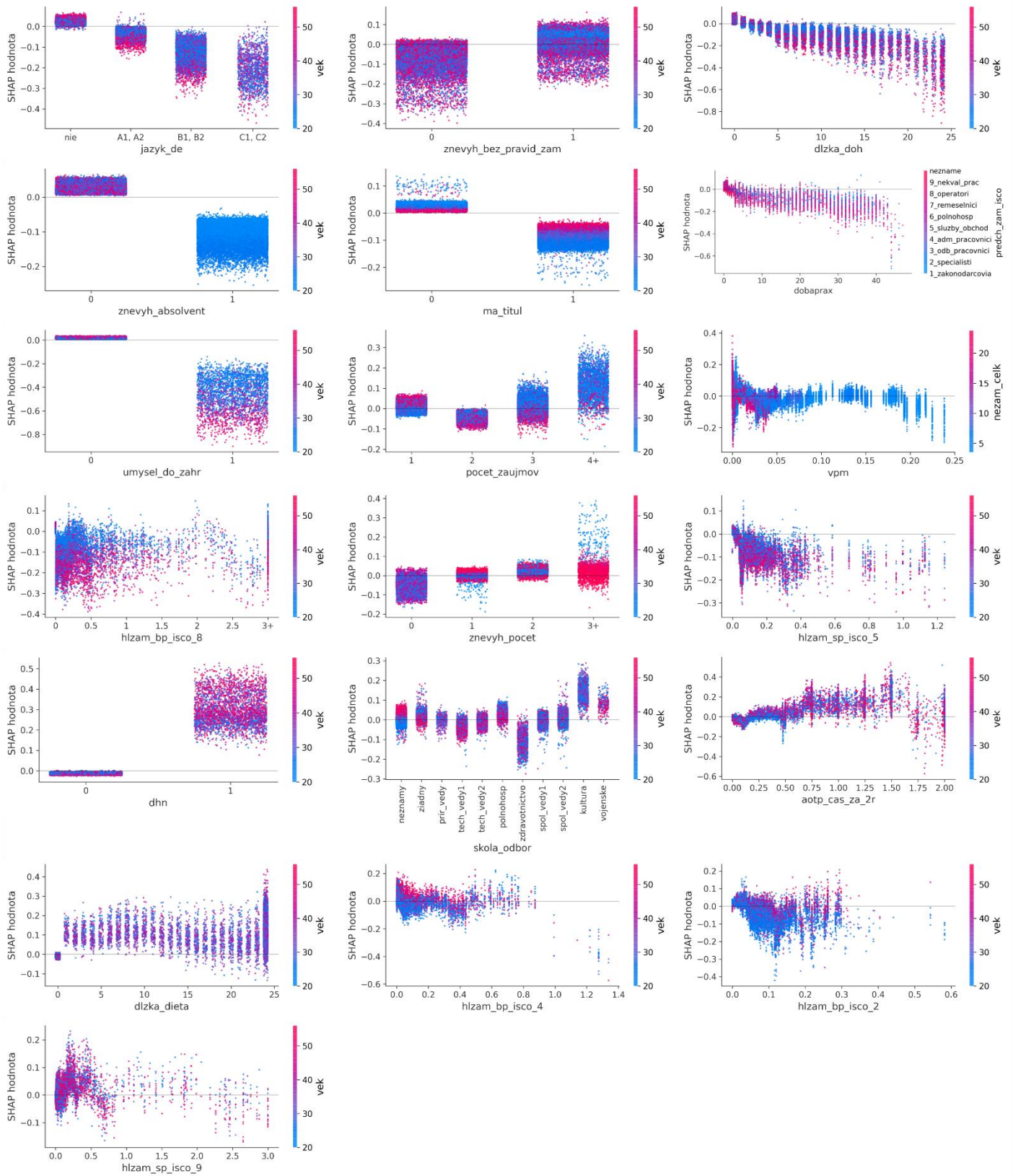
32	hlzam_sp_isco_6	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 6 podľa SK ISCO-08 v prípade záujmu a praxe	priem. 0,132	99,32%
33	hlzam_sp_isco_7	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 7 podľa SK ISCO-08 v prípade záujmu a praxe	priem. 0,244	88,63%
34	hlzam_sp_isco_8	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 8 podľa SK ISCO-08 v prípade záujmu a praxe	priem. 0,401	87,79 %
35	hlzam_bp_isco_1	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 1 podľa SK ISCO-08 v prípade záujmu a bez praxe	priem. 0,073	97,25 %
36	hlzam_bp_isco_3	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 3 podľa SK ISCO-08 v prípade záujmu a bez praxe	priem. 0,068	87,71 %
37	hlzam_bp_isco_5	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 5 podľa SK ISCO-08 v prípade záujmu a bez praxe	priem. 0,105	81,62 %
38	hlzam_bp_isco_6	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 6 podľa SK ISCO-08 v prípade záujmu a bez praxe	priem. 0,147	99,40 %
39	hlzam_bp_isco_7	počet voľných pracovných miest v okrese na 1 UoZ hľadajúceho zamestnanie v hlavnej triede 7 podľa SK ISCO-08 v prípade záujmu a bez praxe	priem. 0,257	88,36 %
40	zdrav_postih	zdravotné postihnutie: nie = 0 / áno = 1	2	0 (98,20 %)
Iné údaje z ÚPSVaR				
41	stup_odk	údaj o stupni odkázanosti: nie = 0 / áno = 1	2	0 (99,29 %)
42	mieraftp	údaj o miere funkčnej poruchy: nie = 0 / áno = 1	2	0 (98,68 %)
43	aotp_ucasti_za_2r	počet účastí na AOTP za 2 roky pred aktuálnym zaradením	priem. 1,33	89,70 %
Údaje zo Sociálnej poisťovne				
44	doch_invalid	poberateľ invalidného dôchodku: nie = 0 / áno = 1	2	0 (97,80 %)
45	doch_iny	poberateľ iného dôchodku (starobný, predčasný starobný, vdovský, vdovecký, sirotsky): nie = 0 / áno = 1	2	0 (99,09 %)
46	vz_szc	mesačný priemerný vymeriavací základ zo SZČ poisťných vzťahov za 2 roky pred aktuálnym zaradením	priem. 443 €	96,93 %
47	vz_dieta	mesačný priemerný vymeriavací základ z materskej dovolenky alebo starostlivosti o dieťa za 2 roky pred aktuálnym zaradením	priem. 508 €	93,62 %
48	vz_ine	mesačný priemerný vymeriavací základ z iných poisťných vzťahov za 2 roky pred aktuálnym zaradením	priem. 480 €	98,16 %
49	dĺzka_szc	počet mesiacov so SZČO poisťným vzťahom za 2 roky pred aktuálnym zaradením	priem. 12,81	96,93 %
50	dĺzka_ine	počet mesiacov s iným poisťným vzťahom za 2 roky pred aktuálnym zaradením	priem. 13,97	98,16 %

Pozn: ^aPre kategorické premenné uvádzame počet kategórií, pre spojité priemer z nenulových hodnôt. ^bPre kategorické premenné uvádzame zastúpenie najpočetnejšej kategórie, pre spojité podiel nulových hodnôt.

B SHAP hodnoty pre ostatné prediktory²⁵



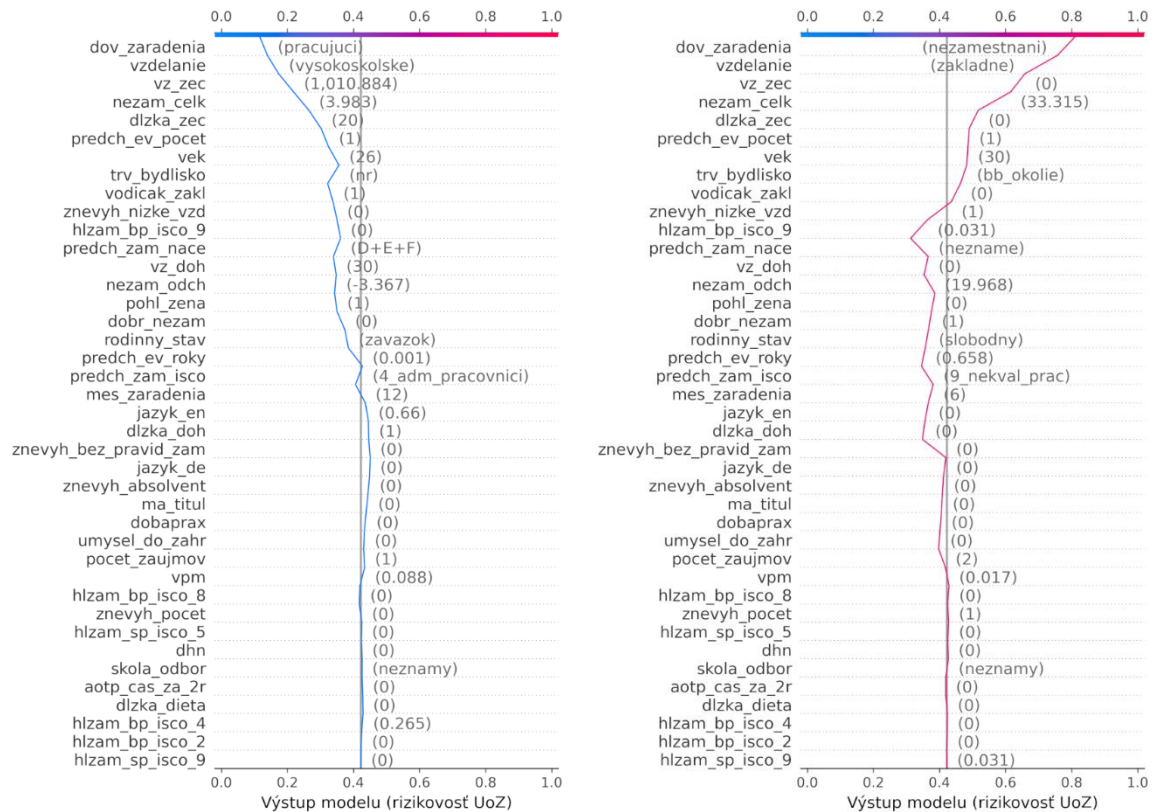
²⁵ SHAP hodnoty pre tri najsilnejšie prediktory zobrazuje Obrázok 18.



Zdroj: vlastné spracovanie pomocou shap.dependence_plot. Pozn.: Každý bod zodpovedá konkrétnemu UoZ. Horizontálna poloha zodpovedá hodnote zvolenej vysvetľujúcej premennej, vertikálna poloha vplyvu tejto hodnoty na rizikovosť. Farba zohľadňuje hodnotu inej zvolenej premennej a pomáha ilustrovať interakcie medzi premennými.

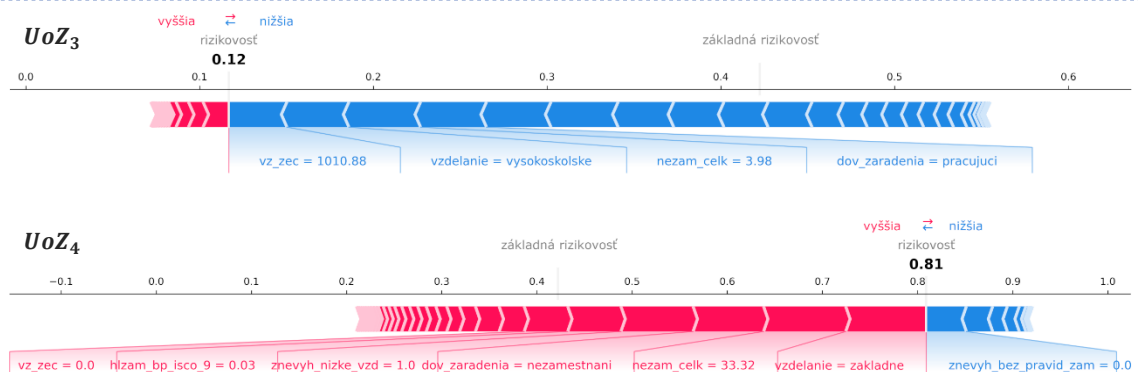
C Modelové príklady nesprávne klasifikovaných uchádzačov

Obrázok 21 Nesprávne klasifikovaní uchádzači s rizikovosťami 11,7 % (UoZ_3 , ľavý panel) a 80,9 % (UoZ_4 , pravý panel)



Zdroj: vlastné spracovanie pomocou shap.decision_plot. Pozn.: Premenné sú zoradené podľa celkovej dôležitosti pre model. V zátvorkách sú zobrazené konkrétne hodnoty premenných pre zvolených UoZ.

Obrázok 22 Najdôležitejšie charakteristiky pre uchádzačov UoZ_3 a UoZ_4 z obrázka 20

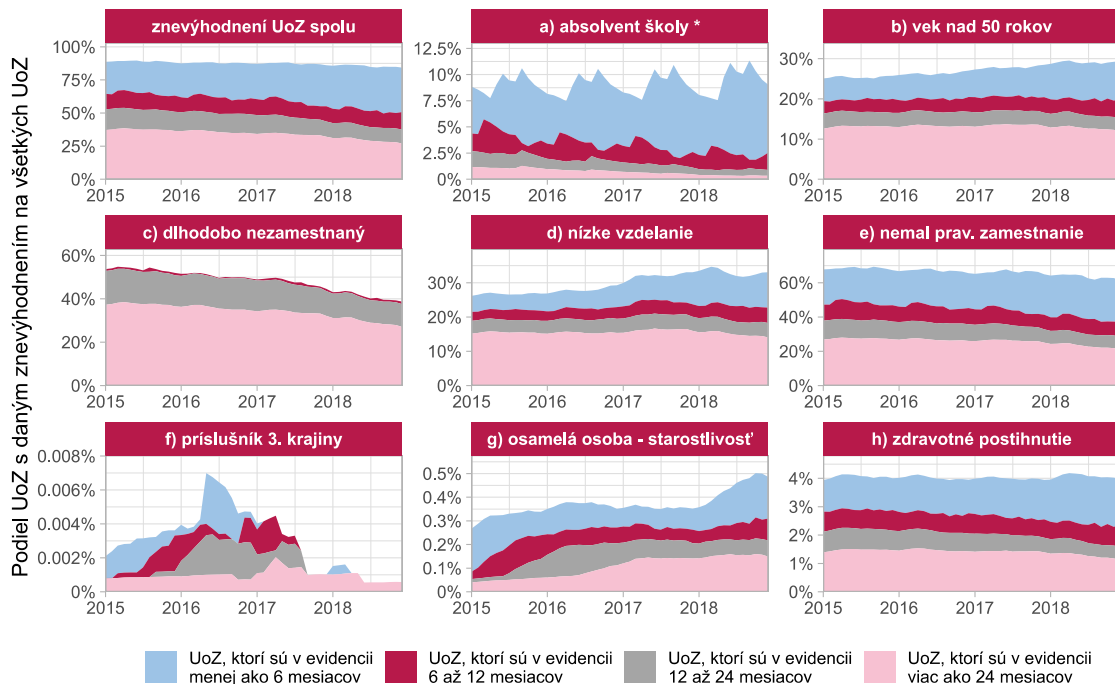


Zdroj: vlastné spracovanie pomocou shap.force_plot. Pozn.: Červenou sú znázornené charakteristiky, ktoré najviac zvyšujú rizikovosť daného uchádzača, modrou charakteristiky, ktoré rizikovosť najviac znižujú.

D Znevýhodnenia a dĺžka evidencie

Zastúpenie rôznych dĺžok nezamestnanosti sa pre jednotlivé typy znevýhodnení líši. Kým pri absolventoch škôl je väčšina vyradená z evidencie skôr ako po šiestich mesiacoch, pri UoZ nad 50 rokov alebo pri UoZ s nízkym vzdelaním je viac ako polovica v evidencii aspoň rok.

Obrazok 23 Stav evidencie UoZ v závislosti od znevýhodnenia a dĺžky pobytu v evidencii



Zdroj: ÚPSVaR. Pozn.: *Pri absolventoch školy uvádzame UoZ, ktorí mali toto znevýhodnenie v čase zaradenia do evidencie, keďže toto znevýhodnenie platí len dva roky od ukončenia sústavnej prípravy na povolanie. Pri ostatných znevýhodneniach uvádzame UoZ, ktorí dané znevýhodnenie mali ku koncu zobrazeného mesiaca.

Podiel UoZ s aspoň jedným znevýhodnením na všetkých evidovaných UoZ v rokoch 2015 až 2018 mierne klesol z takmer 90 % pod 83 %. Pokles v skupinách UoZ, ktorých nezamestnanosť trvá 12 až 24 mesiacov, resp. viac ako 24 mesiacov, bol pritom čiastočne kompenzovaný nárastom medzi uchádzačmi evidovanými menej ako 6 mesiacov.

Na rozdiel od iných znevýhodnení absolventi škôl bývajú nezamestnaní prevažne krátkodobo. Najpočetnejšou skupinou sú absolventi vyradení do 6 mesiacov, naopak len zanedbateľná časť absolventov zostáva v evidencii viac ako 2 roky.²⁶ Dlhodobejší problém s uplatnením sa dá pozorovať pri absolventoch zaradených v septembri, zatiaľ čo medzi absolventmi nezamestnanými menej ako 6 mesiacov pozorujeme zjavný nárast aj v júni. Vzhľadom na inštitucionálne nastavenie²⁷ možno júnový prírastok pripísať prevažne absolventom vysokých škôl a septembrový absolventom stredných škôl. Ako viac problematické sa teda javí uplatnenie absolventov stredných škôl.

Nárast podielu znevýhodnených UoZ na všetkých evidovaných uchádzačoch pozorujeme pri znevýhodneniach nízke vzdelanie a vek nad 50 rokov. Tento nárast sa týka prevažne uchádzačov, ktorých

²⁶ Znevýhodnenie absolvent školy je definované ukončením sústavnej prípravy na povolanie pred menej ako dvomi rokmi (Tabuľka 2). Z toho dôvodu uchádzačovi zaniká toto znevýhodnenie po maximálne dvoch rokoch od zaradenia do evidencie aj v prípade, že je naďalej evidovaný. Aby sme zohľadnili aj UoZ, ktorí boli zaradení s týmto znevýhodnením a ich nezamestnanosť trvá dlhšie, príznak znevýhodnenia absolvent školy nesledujeme k aktuálnemu mesiacu, ale k dátumu zaradenia do evidencie.

²⁷ Zdravotné poistenie je hradené štátom za absolventa strednej školy do konca augusta a za absolventa vysokej školy iba do dňa ukončenia štúdia.

nezamestnanosť trvá menej ako 6 mesiacov. Naopak podiel UoZ nezamestnaných viac ako 2 roky sa medzi takto znevýhodnenými uchádzačmi výrazne nemení.

V rokoch 2015 až 2018 sme zaznamenali pokles zastúpenia dlhodobo nezamestnaných UoZ o 15 percentuálnych bodov z 53,7 % na 38,7 %. Obe skupiny UoZ, ktorých evidencia trvala viac ako 1 rok, sa podieľali na celkovom poklese takmer rovnomerným dielom, keďže zastúpenie UoZ v oboch skupinách pokleslo o približne 30 %. Percentuálny podiel dlhodobo nezamestnaných UoZ klesá aj v prípade UoZ znevýhodnených zdravotným postihnutím, či takých, ktorí pred zaradením nemali pravidelne platené zamestnanie. Pri zdravotne postihnutých je pokles dlhodobo nezamestnaných kompenzovaný nárastom takých UoZ, ktorých nezamestnanosť trvá menej ako 6 mesiacov, takže zdravotne postihnutí stabilne tvoria asi 4 % všetkých UoZ. V prípade UoZ bez predchádzajúceho pravidelného plateného zamestnania sa pokles podielu dlhodobo nezamestnaných premietol aj do celkového poklesu podielu takto znevýhodnených uchádzačov na všetkých UoZ (zo 68 % na začiatku roku 2015 na 61 % na konci roku 2018).