

Data Science Akademie - Intermediate

Posuňte své analytické schopnosti na další level

Máte analytický „mindset“? Už nyní ve své práci často pracujete s daty, ať už v rámci nějaké SQL databáze, v Excelu nebo v Tableau? **Rádi byste se posunuli dál a začali pracovat na komplexnějších projektech?** Máte za sebou i řadu on-line školení na DataCampu nebo na Udemy, kde jste se seznámili s řadou nástrojů jako je R či Python a vyzkoušeli jste si i aplikaci různých ML algoritmů na jednoduchých ukázkových úlohách? A přesto je pro vás obtížné si představit, jak byste uchopili a dodali řešení nějakého reálného byznysového problému? Pokud pro vás platí většina z výše uvedeného, pak je Data Science Akademie přímo pro vás.

V průběhu 8 týdnů budeme společně řešit reálný byznysový problém z oblasti credit risku a to způsobem **end-to-end**. Naučíte se při tom využívat nástroje SQL, Python, R a Tableau k řešení všech dílčích úkolů, které datový vědec typicky řeší v rámci jednotlivých fází projektu - od porozumění byznysovému problému a porozumění datům přes přípravu dat a jejich modelování až po posouzení kvality modelů a jejich nasazení do produkce.

Na rozdíl od on-line kurzů, které se soustředí na výuku dílčích znalostí a dovedností, se v DS Akademii naučíte, jak všechny tyto poznatky využít v rámci jednotného workflow při řešení reálného problému. Velkou výhodou DS Akademie je i možnost průběžné **osobní konzultace s lektory**. Díky tomuto projektovému a osobnímu přístupu se budete moct zapojit do práce na datově-analytických projektech ve vaší vlastní firmě bezprostředně po skončení akademie, případně už v jejím průběhu.

Jak to bude probíhat?

- Na začátku si na základě informací od byznysu vydefinujeme cíle projektu a kritéria posouzení jeho ne/úspěšnosti.
- V SQL databázi si připravíme data a napočítáme prediktory.
- Ukážeme si, jak s pomocí indexování, SQL kódovacích standardů a best practices zajistit, aby naše dotazování bylo optimální z hlediska náročnosti na zpracování a abychom svému SQL kódu rozuměli i za rok, až se k němu budeme vracet.
- Využijeme knihovny z R (*dplyr*, *tidyr*, *ggplot2*, *DataExplorer*) a z Pythonu (*pandas*, *matplotlib*, *seaborn*) k explorační analýze dat.
- Za použití různých knihoven z R a Pythonu (např. *stringr*, *forcats*, *lubridate*, *recipes*, *caret*, *smbinning*, *numpy*, *pandas*, *scikit-learn*) si v rámci feature engineeringu vytvoříme nové, potenciálně užitečné proměnné a na všech datech provedeme potřebné transformace jako je např. imputace chybějících hodnot, binning, one hot encoding, redukce dimenzí, normalizace apod.
- S pomocí ML knihovny *scikit-learn* v Pythonu se pokusíme v našich datech identifikovat skupiny zákazníků s podobným profilem, které potom marketingové oddělení bude moct oslovit cílenou nabídkou finančních služeb.
- S pomocí ML knihovny *caret* v R se pokusíme identifikovat klíčové prediktory rizikových půjček a předpovídat schopnost splácet půjčku u jednotlivých žadatelů o úvěr.
- V R i v Pythonu se naučíme využívat cross validizaci k ladění hyperparametrů použitých modelů a resampling k ošetření situací, kdy existuje značný nepoměr mezi zástupci jednotlivých predikovaných kategorií.
- Ukážeme si, jak posoudit, zda jsou námi vytvořené datové modely kvalitní, a to ze statistického i byznysového hlediska.
- Pokusíme se zlepšit naše výkonové metriky použitím ensemble modelu kombinujícího výstupy většího počtu různých druhů modelů.

- Nasadíme naše modely do produkce, kde budou moct byznysu pomáhat lépe se rozhodovat, a v *R Markdownu*, případně v *Jupyter Notebooku* vytvoříme snadno reprodukovatelný report o projektu pro naše kolegy.
- Využijeme Tableau k vizualizaci úvěrových dat, která umožní lidem z byznysu získat lepší přehled o stavu některých byznysově relevantních metrik a provádět na nich vlastní explorační analýzu.

Co si ze školení (mimo jiné) odnesete?

- Získáte řadu klíčových znalostí a dovedností, které vám umožní realizovat datově-analytické projekty střední úrovně komplexity způsobem end-to-end.
- V SQL databázi si budete umět připravit a předzpracovat data, která budete potřebovat pro modelování zadaného byznysového problému.
- S pomocí programovacích jazyků Python a R se naučíte rychle se zorientovat v datech, posoudit jejich relevanci pro zadaný byznysový problém a provést potřebné datové transformace.
- S pomocí stejných nástrojů dokážete aplikovat klastrovou analýzu a vybrané binární klasifikátory na odpovídající byznysové problémy a budete umět posoudit kvalitu výsledných datových modelů ze statistického i byznysového hlediska.
- Budete znát postupy, jak odladit hyperparametry datových modelů a jak ošetřit výrazný nepoměr predikovaných kategorií, a naučíte se také při modelování dat využívat principu ensemblingu.
- Budete vědět, jak vytvořené modely nasadit v byznysové praxi a jak je použít na nová data. Naučíte se také vytvořit snadno reprodukovatelný report o Vašem projektu.
- Budete znát základní principy vytváření užitečných, prakticky použitelných a zároveň esteticky přitažlivých datových vizualizací v Tableau.

Co byste už měli znát a umět?

D-S Akademie je primárně určena pro ty, kteří se chtějí v oblasti data science posunout z úrovně začátečník na úroveň středně pokročilý/á. **Pro absolvování kurzu jsou nutné základy zejména v oblasti SQL**, ve které se kurz zaměřuje pouze na pokročilejší příkazy. Svou úroveň si otestujte v jednoduchém testu, který lze [stáhnout zde](#). Rádi Vám doporučíme, jak si případně znalosti před začátkem kurzu doplnit.

Společnost Deloitte je předním globálním poskytovatelem služeb v oblasti auditu a assurance, podnikového poradenství, finančního poradenství, poradenství v oblasti rizik a daní a souvisejících služeb. Naše síť členských firem ve více než 150 zemích a teritoriích poskytuje služby čtyřem z pěti společností figurujících v žebříčku Fortune Global 500®. Chcete-li se dozvědět více o způsobu, jakým zhruba 264 000 odborníků dělá to, co má pro klienty smysl, navštivte www.deloitte.com.

Společnost Deloitte ve střední Evropě je regionální organizací subjektů sdružených ve společnosti Deloitte Central Europe Holdings Limited, která je členskou firmou sdružení Deloitte Touche Tohmatsu Limited ve střední Evropě. Odborné služby poskytují dceřiné a přidružené podniky společnosti Deloitte Central Europe Holdings Limited, které jsou samostatnými a nezávislými právními subjekty. Dceřiné a přidružené podniky společnosti Deloitte Central Europe Holdings Limited patří ve středoevropském regionu k předním firmám poskytujícím služby prostřednictvím více než 6 000 zaměstnanců ze 44 pracovišť v 18 zemích.

Tato publikace obsahuje pouze obecné informace a společnost Deloitte Touche Tohmatsu Limited ani žádná z jejích členských firem či jejich spřízněných podniků (souhrnně „síť společností Deloitte“) jejím prostřednictvím neposkytuje odborné rady a služby. Přijetí jakéhokoliv rozhodnutí či jednání, které může mít dopad na Vaše finance či podnik, byste měli konzultovat s kvalifikovaným odborným poradcem. Žádný subjekt v rámci sítě společností Deloitte nenes odpovědnost za ztráty vzniklé jakýmkoli osobám v důsledku použití této komunikace.

„Deloitte“ nebo „DTTL“ označuje jednu či více společností Deloitte Touche Tohmatsu Limited, britské privátní společnosti s ručením omezeným zárukou, jejích členských firem a jejich spojených osob. Deloitte Touche Tohmatsu Limited ani žádná z jejích členských firem nenesou odpovědnost za konání či pochybení ostatních členských firem. Každá členská firma je samostatným a nezávislým právním subjektem, který působí pod názvem „Deloitte“, „Deloitte & Touche“, „Deloitte Touche Tohmatsu“ či jiným obdobným názvem. „Deloitte ve střední Evropě“, „DCE“, „firma“ nebo „my“ označuje jeden nebo více subjektů sdružených ve společnosti Deloitte Central Europe Holdings Limited, která je členskou firmou sdružení Deloitte Touche Tohmatsu Limited ve střední Evropě. Odborné služby poskytují dceřiné a přidružené podniky společnosti Deloitte Central Europe Holdings Limited, které jsou samostatnými a nezávislými právními subjekty. Společnost Deloitte Advisory s.r.o. je dceřinou společností Deloitte Central Europe Holdings Limited.

© 2019 Pro více informací kontaktujte Deloitte Česká republika.