

Klasifikace aut podle místa výroby

Cílem tohoto úkolu je nalézt model, který bude klasifikovat auta podle místa výroby a odhadnout chování tohoto modelu na reálných datech.

Požadované kroky analýzy:

1. Vyberte alespoň dva klasifikátory vhodné pro tento typ dat. Svůj výběr zdůvodněte? (2 bod)
2. Na vhodné podmnožině dat vybrané klasifikátory natrénujte. Jaké úspěšnosti dosahují? Podle jakých příznaků se rozhodují? Dává to smysl? (5bodů)
3. Jakou úspěšnost klasifikace očekáváte v hypotetickém reálném nasazení Vašeho klasifikátoru, tj. v případě klasifikace dalšího auta (které nemáte v datovém souboru, ale u kterého máte naměřeny příslušné příznaky)? (1 bod)
4. Za předpokladu, že výběr aut v datovém souboru je reprezentativní, odhadněte pravděpodobnost, že další klasifikované auto bylo vyrobeno v USA. (1 bod)
5. Jaká je pravděpodobnost, že další klasifikované auto, ač vyrobené v USA, bude klasifikováno jako vyrobené mimo USA? Jaká bude naproti tomu pravděpodobnost, že další klasifikované auto, vyrobené mimo USA, bude klasifikováno jako vyrobené v USA? Kterou chybu považujete za závažnější a proč? (3 body)
6. Má ve Vašem případě na přesnost klasifikace vliv to, zda je trénovací (testovací) množina vyvážená? Pokud ano, jaký? Vyvažovali jste trénovací (testovací) množinu? Pokud ano, proč a jak? Pokud ne, proč ne? (5 bodů)
7. Shrňte vlastnosti vybraných klasifikátorů a vyberte nejvhodnější klasifikátor pro tuto úlohu. (3 body)

Výsledky upravte do formy zprávy, která bude obsahovat stručný **úvod**, popis metod, které jste použili, v sekci **metody**, výsledky jejich aplikace na data v sekci **výsledky** a závěry, které jste zjistili interpretací výsledků v sekci **závěr**. Maximální délka zprávy je 5 stran. Zprávu ve formátu pdf odevzdejte pomocí UploadSystemu.