Okruhy, ze kterých si vylosujete teoretickou otázku ke zkoušce `M8DM1 Data mining I`:

1. Data mining, KDD proces a vše kolem.
2. Dataminingové metodologie.
3. Relační databáze a jazyk SQL.
4. BI - datový sklad, OLAP, OLTP.
5. Integrace dat.
6. Čištění dat.
7. Transformace dat.
8. Redukce dat.
9. Praktická aplikace modelu na data.
10. Praktická interpretace modelu a jeho parametrů.

Matematické (metodologické) otázky ke zkoušce `M8DM1 Data mining I`:

1. **Analýza hlavních komponent**. Popište cíle analýzy hlavních komponent. Jak jsou komponenty konstruovány? V čem spočívá redukce dimenze? Jak se interpretují její výsledky? Jak se v praxi aplikuje?

2. **Faktorová analýza**. Popište cíle a model faktorové analýzy. Jak se faktory hledají? Jak se v praxi aplikuje? Jak se interpretují její výsledky? K čemu slouží rotace?

3. **Mnohorozměrné škálování**. Popište úlohu mnohorozměrného škálování. Popište algoritmus metrického škálování a zobrazení dat v prostoru nízké dimenze. Jaký je rozdíl mezi metrickým a nemetrickým škálováním?

4. **Exploratorní analýza dat**. Popište k čemu slouží exploratorní analýza dat. Popište metody jednorozměrné a mnohorozměrné exploratorní analýzy pro numerické i kategoriální proměnné. Vizualizace dat.

5. **Kontingenční tabulky**. Popište testy nezávislosti v kontingenčních tabulkách. Jak se v nich měří závislosti? Popište znaménkové schéma. K čemu se používá? Co to je a k čemu se používá korespondenční analýza? Popište její základní myšlenky. Jak se interpretují její výsledky?

6. **Analýza nákupního košíku**. Popište analýzu nákupního košíku. Jaké číselné charakteristiky pravidel se používají? Jak se hledají pravidla pro dvou i víceprvkové množiny? Popište její zobecnění pro negované položky a hierarchické struktury dat.

7. **Shluková analýza**. Popište úlohu shlukové analýzy. Popište algoritmus a uveďte metody hierarchického shlukování. V čem se nehierarchické shlukování liší od hierarchického. Popište metodu $k$-means a $k$-medoids. Jaké metody se používají pro určení výsledného počtu shluků?

8. **Lineární regrese**. Popište model lineární regrese, jeho předpoklady a interpretujte parametry modelu. Co to je multikolinearita? Jak se identifikuje a jaké může mít následky? Popište hřebenovou regresi a LASSO. K čemu se tyto metody používají?

9. **Logistická regrese**. Popište model logistické regrese. Co znamenají jednotlivé parametry tohoto modelu? Co to je logistické skóre? Jak se v logistické regresi odhadují hodnoty závisle proměnné? Co to je ROC a Lorenzova křivka? Uveďte číselné charakteristiky odvozené od těchto křivek.

10. **Rozhodovací stromy**. Jakou úlohu řešíme pomocí rozhodovacích stromů? Popište algoritmy CART a CHAID. K čemu slouží a jak funguje prořezávání? Uveďte číselné charakteristiky popisující kvalitu modelu.