

Er waren mogelijks 2 reeksen maar stond er volgens mij niet bij welke reeks het was (Dit is de reeks met het witte voorblad ipv het groene) Examen 12/06

10 MC

Vraag over LOF en het onderliggende principe

Vraag over closed sequence pattern and what it is.

- bij die vraag stond er iets van subsumed (heel gelijkaardig als het voorbeeldexamen (superseded) maar idk of t zelfde is

Vraag over GMM in vergelijking met DBSCAN

Vraag over closure

Vraag met allemaal stelling:

- Een ging over dat boundaries van SVMs en of ze lineair zijn
- iets over non linearity in decision trees

Welke metric is cut off dependent :

- lift
- AUC
- AUCPR
- F1

Vraag over wat de alpha in cost complexity pruning is

Open Vraag 1:

One of the main challenges of fraud detection is class imbalance. Explain what it exactly entails and why it can be a challenge. What are possible techniques you can use to deal with these challenge and explain how these influence the creation of deployment of the model (idk wat de laatste zin betekende of wat er exact stond)

Open vraag 2:

Compare kNN, Naive Bayes and C4.5 in the terms of incremental/ lazy learner. Include in the discussion a comparison in terms of prediction efficiency, adaptability to data changes, model creation, training time.

Open vraag 3:

Data Leakage: What is it, what are the two most important types of data leakage and provide an example of each type. Include in the comparison of both types an evaluation of their impact on model performance & reliability. Discuss impact on accuracy, possible biases and how these risks can be mitigated

Open vraag 4:

Regularization: What is regularization and what is it used for. Explain what different types of regularization exist and compare their advantages / disadvantages. Probably there was still another subquestion, ik denk iets van invloed op andere fases van development proces.