**NARRATIVE FOR MILESTONE 4: DATABASE ENHANCEMENT**

**Table of Contents**

[Artifact Description 3](#_Toc190474117)

[Justification for Inclusion 3](#_Toc190474118)

[Meeting Course Outcomes 3](#_Toc190474119)

[Reflection on the Enhancement Process 4](#_Toc190474120)

[Conclusion 4](#_Toc190474121)

[Bibliography 5](#_Toc190474122)

# Artifact Description

The artefact is the Python script for network intrusion detection using the UCI KDD Cup 1999 dataset that has been enhanced. The script as an initial version of itself had a machine learning pipeline for preprocessing, feature transformation, model training and evaluation. The enhancement of this increases it by making it possible to store and query processed data efficiently by integrating database functionality to do so. It was created in the frame of the CS 499 Capstone Project, and partly expanded on work, that had been done earlier in courses of the Computer Science program.

# Justification for Inclusion

This artifact also demonstrates that I was able to bring together database solutions with a real-world machine learning application which I have included in my ePortfolio. The enhancements showcase my skills in database design, data integrity, query optimization, and visualization. Specific improvements include:

* Design a normalized database schema for the storage of features as well as labels separately.
* Visualizations are added such as class distribution and one visualization example features importance plots.
* Errors handling for robust database operation.
* Training and evaluating Multiple classifiers( KNN, Logistic Regression, Naive Bayes, Decision Tree, SVM) and comparing their accuracy score.

They show that I can remedy these things to handle real-world problems with scalable, efficient and interpretative solutions.

# Meeting Course Outcomes

This enhancement aligns with several Computer Science program outcomes:

1. **Databases:** Designed and implemented a normalized database schema for efficient data storage and retrieval.
2. **Software Engineering Best Practices:** Applied modular programming, error handling, and visualization techniques.
3. **Performance Optimization:** Optimized database queries and feature selection to improve model efficiency.
4. **Integration:** Integrated database functionality into a machine learning pipeline.

I met the outcomes I planned to achieve in Module One, particularly in demonstrating proficiency in database design and integration. Moving forward, I plan to focus more on database security and scalability to further enhance my skills.

# Reflection on the Enhancement Process

Through this, I learned how to make an integration between databases and machine learning workflows and taught myself a quick course on database schema design and query optimisation on the way. I also enhanced my skill at using visualizations to read data and model performance as a route to convey insights.

The insertion of large datasets into the database was one of the main challenges that I had to face. Debugging SQL queries for retrieving complex data also was troublesome but that helped me to comprehend the very core database systems better. I did manage to solve the problem further, by balancing model complexity vs. interpretability to further enhance my ability to problem solve.

This project significantly contributed to my professional growth, helping me design and implement scalable solutions for real-world problems. I am happy with the skills I learned, and I am ready to use them in future projects.

# Conclusion

This enhancement was a significant step in my computer science journey, allowing me to integrate database functionality into a machine learning pipeline. The script has become more robust, scalable and useful for real-world deployment. I believe that all the skills I acquired through this project will be very beneficial for my future profession.

# 

# Bibliography

Qiao, S. and Shen, C., 2021. From Text-based Genome, Population Variations, and Transcriptome Datafiles to SQLite Database and Web Application: A Bioinformatical Study on Alfalfa. *bioRxiv*, pp.2021-08.

Suraya, S. and Sholeh, M., 2022. Designing and implementing a database for thesis data management by using the python Flask Framework. *International Journal of Engineering, Science and Information Technology*, *2*(1), pp.9-14.

Veit, M., Sørensen, C.G., Pomianowski, M.Z., Johra, H. and Marszal-Pomianowska, A., 2023. A practical approach to set up a simple database architecture using SQLite and Python.

Verma, A. and Sharma, B., 2022. Dynamic E-Certificate Designing with Automatic Mailing System using Python and SQLite3. *Journal of Applied Information Science*, *10*(2).