Big Project 2 IF3170 - Artificial Intelligence 2024/2025

Group Name:cetjipiti
Members:

- Jimly Nur Arif (Student ID: 13522123])
- Yosef Rafael Joshua (Student ID: 13522133)
- Samy Muhammad Haikal (Student ID: 13522151)
- Muhammad Roihan (Student ID: 13522152)

Table of Contents

---

1. Introduction

Provide an overview of the project's objectives, the importance of machine learning, and the significance of applying these algorithms to the UNSW-NB15 dataset.

---

2. Dataset Description

Detail the UNSW-NB15 dataset, including:

- Overview of the dataset (network traffic data, attack categories, etc.)
- Variables and their descriptions
- Link to dataset source

---

3. Algorithm Implementation

3.1 K-Nearest Neighbors (KNN)

- Description of the algorithm
- Steps taken for implementation from scratch
- Parameters supported (e.g., distance metrics, number of neighbors)

3.2 Gaussian Naive Bayes

- Description of the algorithm
- Implementation details and assumptions

3.3 ID3 Algorithm

- Explanation of ID3 (decision tree building)
- Handling of numerical data

---

4. Data Preprocessing

Explain the data cleaning and preprocessing steps, such as:

- Handling missing values
- Encoding categorical variables
- Normalization/Standardization
- Feature selection and dimensionality reduction (if any)

---

5. Experiments and Results

Describe the experimental setup:

- Train-test split or cross-validation approach
- Performance metrics used (e.g., accuracy, precision, recall, F1-score)

Present the results of:

- Implementations from scratch
- Comparisons with scikit-learn library

Use visualizations (tables/graphs) to support findings.

---

6. Insights and Discussion

Discuss key observations, including:

- Differences between from-scratch implementations and scikit-learn results
- Challenges faced during implementation
- Strengths and weaknesses of each algorithm

---

7. Conclusion and Recommendations

Summarize findings and provide recommendations for future work.

---

8. References

Provide all references, including links to the dataset, scientific articles, and other resources used.