DALHOUSIE
UNIVERSITY

Machine Learning for Network Intrusion Detection Systems

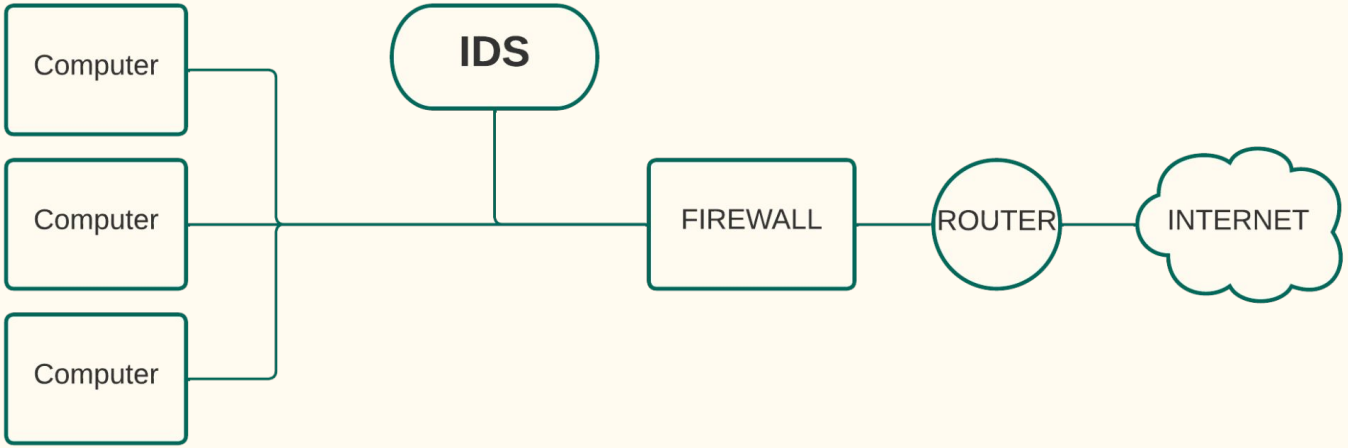
—

Group 13

Corentin Goetghebeur ~ Gabriel Marchand ~ Rinchen Toh

Problem Statement - Intrusion Detection System

- Network Traffic Analysis
- Reporting Dangerous Behaviors



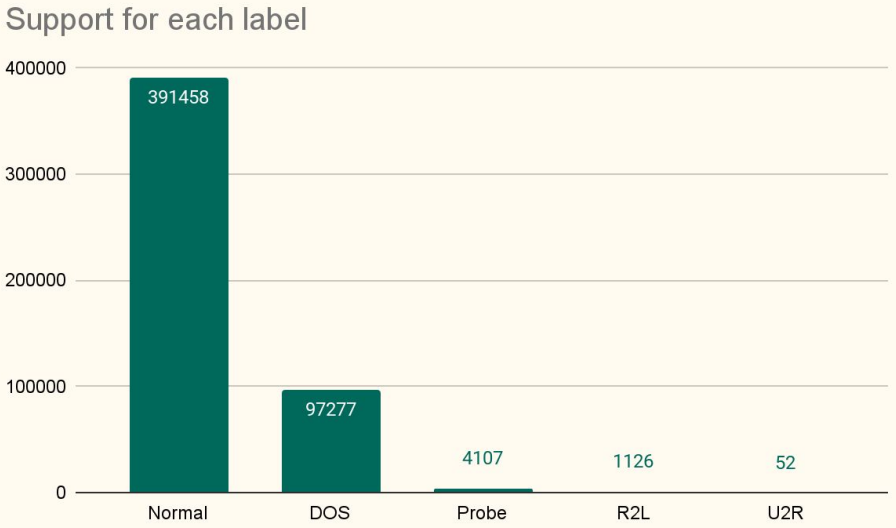
What We Want to Achieve

- To compare the effectiveness of different Machine Learning algorithms in detecting and hence prevent possible fraudulent cyber attacks on the network systems,

- We aim at finding the best model to achieve this goal

Data Set

- KDD cup'99 dataset
  - 42 Columns & around 500,000 Rows
  - Network traffic
  - Widely used for research
- Uneven support for labels
  - Grouping together



Methodology

Pre-Processing | Implementation | Evaluation

- Feature Mapping
- Correlation
- Grouping Labels
- Train-Test Split

- One notebook per algorithm
- Scaling
- Training

- Evaluation
- Grouping
- Comparing

Proposed Solutions

- Naive Bayes
- SVM
- Logistic Regression
- Random Forest
- KNN
- Decision Tree

Evaluation Metrics

Classification Report

All information

Details for each label

+

Confusion Matrix

Detailed prediction results

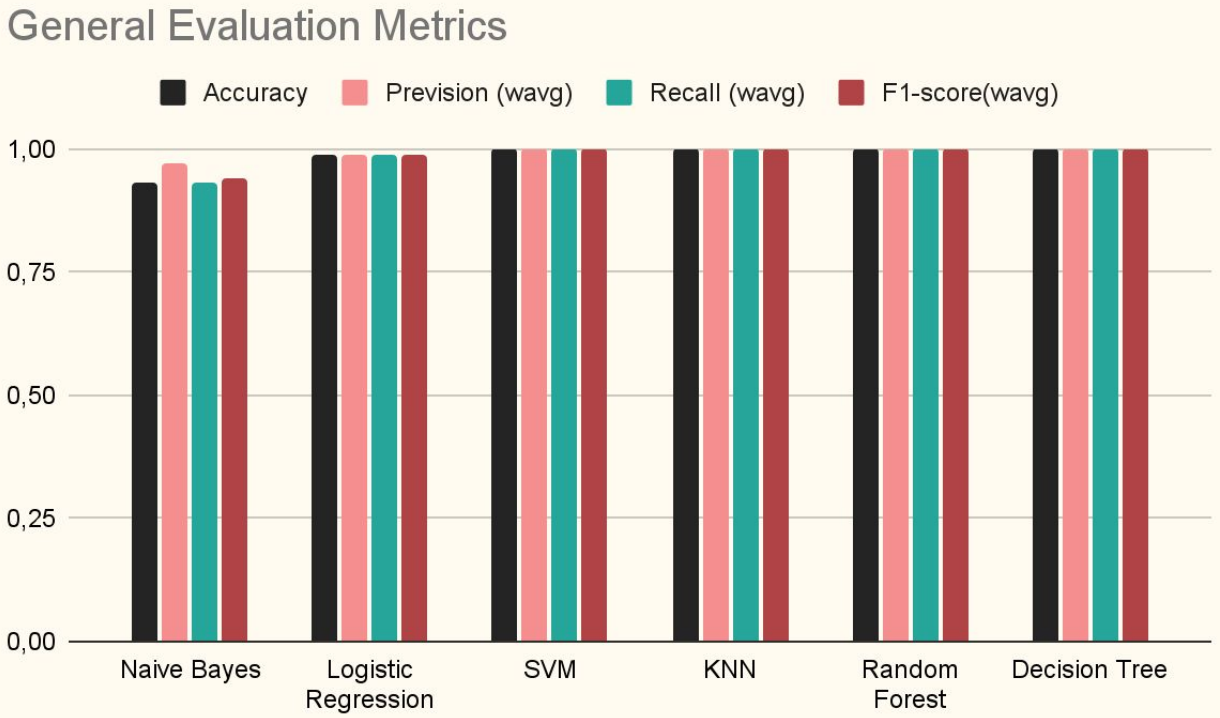
→

Recall

Identifying all the dangerous rows

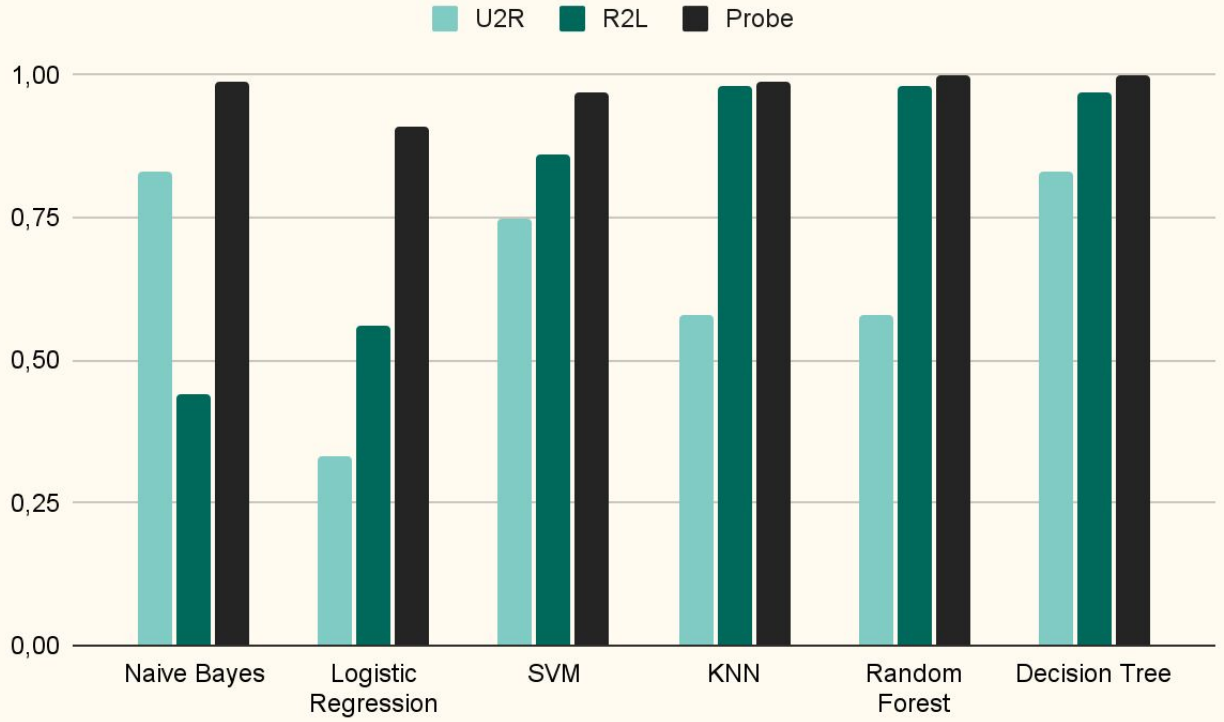
Most important: *U2R* and *R2L*

Results - Overall prediction scores



(*wavg = weighted average)

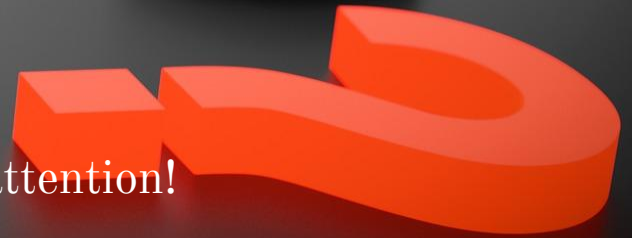
Results - Recall on the most impactful labels



Future Scope

- **Cleaned and more recent versions of our dataset — KDD CUP'99**
- **More sophisticated ML algorithms (e.g. J48 Classifier)**
- **Deep Neural Network for better calculation on large datasets**

Conclusion



Thank you for your attention!

Contact - Group 13

- Corentin Goetghebeur ~ cr453043@dal.ca
- Gabriel Marchand ~ gb614643@dal.ca
- Rinchen Toh ~ rn835427@dal.ca

- Code: https://github.com/CorentinGoet/ML_IDS