IEEE-CIS Fraud Detection Challenge

Comparative Study of SVM and Decision Tree Binary Classification

Khashayar Zardoui
Dept. Computer Science & Software Engineering
Concordia University
Montreal, Canada
khashayar.zardoui@mail.concordia.ca
ID: 40052568

Paolo Junior Angeloni
Dept. Computer Science & Software Engineering
Concordia University
Montreal, Canada
p_ange@live.concordia.ca
ID: 25976944

***Abstract*—abstract: This report evaluates the performance...**

I. The Fraud-Detection Pipeline

...

1) Data loading and exploration (EDA)
2) Removal, Imputation, Label Encoding and Scaling
3) Model training and performance assessment

II. Exploratory Data Analysis

A. Data Structure Inspection

1) Before any data transformation, we observed the `train` and `test` datasets had a mixture of `float64`, `int64` and `object` types
2) missing values
   description goes here...
3) target balance
   description goes here...

B. Statistical Summary & Visualizations

Fig. 1. some image here

TABLE I
SOME STATS...

Metric	Value
one	...
two (%)	...
three	...
four	...

C. Findings & Hypotheses

...

...

III. Data Pre-Processing & Cleaning

A. Imputation & Removal

1) ...
2) ...
3) ...

B. Normalize & Scale Features

1) ...
2) ...
3) ...

C. Encoding Categorical Features

1) ...
2) ...
3) ...

IV. Models

intro to the models used

A. Support Vector Machine (SVM) Classifier

Due to the size of the dataset (590,000+ samples), a standard SVM with a non-linear kernel ($O(n^3)$) was computationally infeasible. We opted for a LinearSVC ($O(n)$) to utilize the entire training set. To satisfy the hyperparameter tuning requirement2, we tuned the Regularization parameter (C) and the Loss function (Hinge vs. Squared Hinge) instead of the kernel.

experiment hyperparameters (C, gamma, kernel etc)

cross-validation and validation splits to evaluate performance

results using different hyperparameters

training and test metrics: confusion matrix, precision, recall, F1-score, and accuracy

B. Decision Tree Classifier

experiment hyperparameters (max depth, min samples split, criterion)
cross-validation and validation splits to evaluate performance
results using different hyperparameters
training and test metrics: confusion matrix, precision, recall, F1-score, an

V. Model Comparison

TABLE II
SOME STATS...

Metric	SVM	Decision Tree
one	...	...
two (%)	...	...
three	...	...
four	...	...

a) discuss similarities & differences. use table:

ACKNOWLEDGMENT

We would like to thank Professor Arash Azarfar and Firat Oncel for their guidance and support throughout this project. Large Language Models, like Google's Gemini were used in an educational context to further understand the resources for this research.

REFERENCES

[1] Numpy, "Numpy API Documentation," [Online]. Available: https://numpy.org/doc/stable/. [Accessed: Nov. 25, 2025].
[2] matplotlib, "Matplotlib API Documentation," [Online]. Available: https://matplotlib.org/stable/index.html. [Accessed: Nov. 25, 2025].
[3] pandas, "pandas API Documentation," [Online]. Available: https://pandas.pydata.org/docs/. [Accessed: Nov. 26, 2025].
[4] seaborn, "seaborn API Documentation," [Online]. Available: https://seaborn.pydata.org/api.html. [Accessed: Nov. 26, 2025].
[5] data preprocessing, "sklearn API Documentation," [Online]. Available: https://scikit-learn.org/stable/modules/preprocessing.html. [Accessed: Nov. 28, 2025].
[6] data imputation, "sklearn API Documentation," [Online]. Available: https://scikit-learn.org/stable/modules/impute.html. [Accessed: Nov. 28, 2025].
[7] label encoding, "sklearn API Documentation," [Online]. Available: https://scikit-learn.org/stable/modules/preprocessing_targets.html#label-encoding. [Accessed: Nov. 28, 2025].
[8] hyperparameter tuning using GridSearchCV, "sklearn API Documentation," [Online]. Available: https://scikit-learn.org/stable/modules/grid_search.html#grid-search. [Accessed: Nov. 30, 2025].
[9] Support Vector Machines, "sklearn API Documentation," [Online]. Available: https://scikit-learn.org/stable/modules/svm.html. [Accessed: Nov. 30, 2025].
[10] Decision Trees , "sklearn API Documentation," [Online]. Available: https://scikit-learn.org/stable/modules/tree.html. [Accessed: Nov. 30, 2025].
[11] XGBoostClassifier, "XGBoost API Documentation," [Online]. Available: https://xgboost.readthedocs.io/en/stable/. [Accessed: Nov. 30, 2025].
[12] metrics, "sklearn API Documentation," [Online]. Available: https://scikit-learn.org/stable/api/sklearn.metrics.html. [Accessed: Dec. 02, 2025].