Fraud detection with AI

Literature review on fraud detection using AI

Credit card fraud detection using AdaBoost and majority voting

K Randhawa, CK Loo, M Seera, CP Lim... - IEEE access, 2018

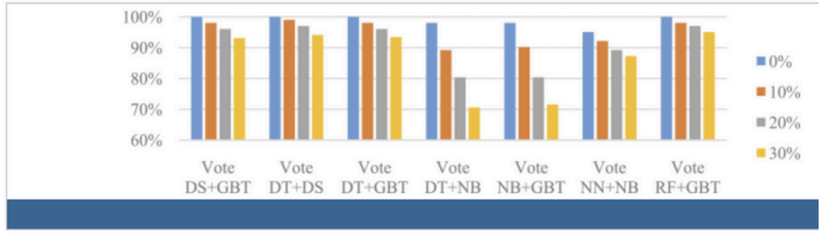


FIGURE 1.
Fraud detection rates with different percentages of noise.

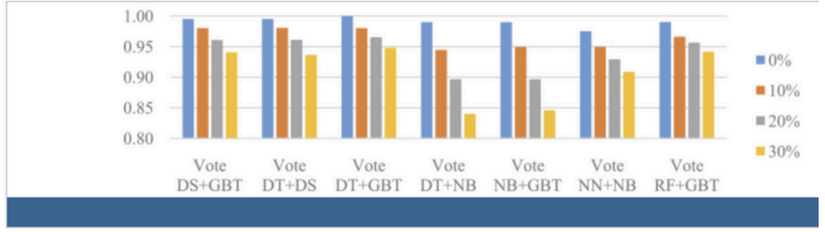
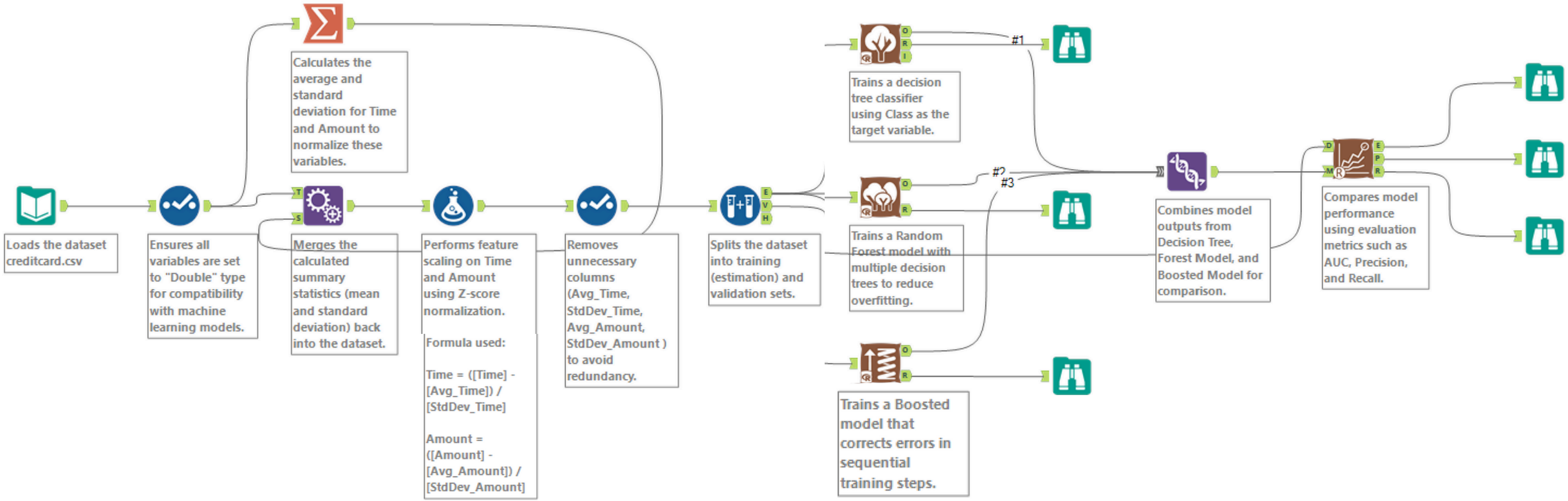


FIGURE 2.
MCC scores with different percentages of noise.

Model	Strengths	Limitations
Bayesian	Good for binary classification problems; efficient use of computational resources; suitable for real-time operations.	Need good understanding of typical and abnormal behaviors for different types of fraud cases
Trees	Easy to understand and implement; the procedures require a low computational power; suitable for real-time operations.	Potential of over-fitting if the training set does not represent the underlying domain information; re-training is required for new types of fraud cases.
Neural Network	Suitable for binary classification problems, and widely used for fraud detection.	Need a high computational power, unsuitable for real-time operations; re-training is required for new types of fraud cases.
Linear Regression	Provide optimal results when the relationship between independent and dependent variables are almost linear.	Sensitive to outliers and limited to numeric values only.
Logistic Regression	Easy to implement, and historically used for fraud detection.	Poor classification performances as compared with other data mining methods.
Support Vector Machine	Able to solve non-linear classification problems; require a low computational power; suitable for real-time operations.	Not easy to process the results due to transformation of the input data.

Presentation of our Alteryx workflow



Results & Model Comparison

Model Comparison Report

Fit and error measures

Model	Accuracy	F1	AUC	Accuracy_0	Accuracy_1
Decision_Tree_Fraud	0.9985	0.6182	0.8465	0.9991	0.6939
Forest_Model_Fraud	0.9991	0.7059	0.9577	0.9998	0.6122
Boosted_Model_Fraud	0.9993	0.7957	0.9859	0.9998	0.7551

Confusion matrix of Boosted_Model_Fraud

	Actual_0	Actual_1
Predicted_0	28425	12
Predicted_1	7	37

Confusion matrix of Decision_Tree_Fraud

	Actual_0	Actual_1
Predicted_0	28405	15
Predicted_1	27	34

Confusion matrix of Forest_Model_Fraud

	Actual_0	Actual_1
Predicted_0	28426	19
Predicted_1	6	30

- The Boosted Model (Boosted_Model_Fraud) is the best-performing model for fraud detection. It has:
- The highest F1-score (0.7957), meaning it balances precision and recall well.
- The highest AUC (0.9859), meaning it distinguishes well between fraud and non-fraud cases.
- The highest Accuracy_1 (0.7551), indicating better detection of fraudulent transactions.
- A relatively low false negative rate (only 7 fraud cases misclassified).

Results & Model Comparison

Model Comparison Report

Fit and error measures

Model	Correlation	RMSE	MAE	MPE	MAPE
Decision_Tree_Fraud	-0.6212	0.6212			-57,924.4898
Forest_Model_Fraud	0.7552	0.0272	0.0014	-2.1371	83.2439
Boosted_Model_Fraud	0.7988	0.0256	0.0011	-11.8982	61.0512

- Boosted Model (XGBoost/AdaBoost) is the most accurate:
  - Highest correlation (0.7988) → Strongest relationship between predictions and actual values.
  - Lowest RMSE (0.0256) and MAE (0.0011) → Most precise predictions.
  - Lowest MAPE (61.05%) → Smallest percentage error.

Thank you