

Anfertigen eines CAT bzw. PIKE

Kann ... // Algorithmus x // exact ... // (Kunden)-Problem ... berechnen / lösen?

Yes, a Ensemble technique can performe the classification

Big Data / Data Mining / Data Science Kernaussage:

(P) roblem):

[Welcher Frage hat für die Lösung des Kunden / Auftraggeber die größte Bedeutung?]:

As we are moving towards the digital world ,cybercrime has been in the rise . Credit card fraud is presently the most frequently occurring problem.Hence there need to detect abnormal activities . This project aims to focus mainly on machine learning algorithms for dectecting fraud.

(I) ntervention:

(Bibliotheken- und Algorithmen-Auswahl, ... z.B. pandas für Finanzdaten ...)

[Welche Berechnung erwäge ich vornehmlich?]:

For data analysis and pre-processing pandas, matplotlib and seaborn. For building the models ensemble technique is used like random forest classifier and XGBoost classifier. Also to find the best parameters ,Hyperparameter tuning methods like RandomizedSearchCV and GridSearchCV methods are implemented . Also for High imbalance in the data Class weight technique is used to give weightage to the minority class.

(K) ontrollintervention

(falls erforderlich: Bliotheken- und Algorithmen-Auswahl ... z.B. scikit-learn für Finanzdaten ...)

[Was ist die andere Möglichkeit?]:

Other machine learning Alogorithms could be used for the fraud detection problem and also many other automated hyperparameter tuning methods can be implemented .There could have been used methods like undersampling or oversampling.

(E) rgebnismaß (Zielgröße(n)) – Die Evidence

[Was möchte ich / der Kunde erreichen? Z.B. Prädiktor oder Klassifikator erstellen ...]:

To identify fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

Die Suche nach der besten Evidenz

1.Problem

To creat an Automated system to detect fraudulent transactions

2.Definition einer wichtigen suchbaren Frage

Does model identifies the fraud transaction?

3.Auswahl der wahrscheinlichsten Quelle für diese Evidenz

After training, the performance of the model is tested on a selection of the existing test data and evaluated.

4.Erstellung einer Suchstrategie

Different Algorithms can be used to achieve the accuracy , but in this project i have used ensemble techniques and due high imbalance in data i have used class weights to give more weightage to minority class .

5.0 Zusammenstellung der Evidenzausbeute

A random forest classifier model and XGBoost classifier model is created for a dataset which contains 284315 samples and 31 Features using best parameters obtained from RandomizedSearchCV and GridSearchCV

6.Anwendung der Evidenz

I have built two models Random Forest Classifier and XGBoost classifier to detect fraud transactions depending the features given.