IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

Anomaly Detection in Credit Card Transaction and Analysis Using Microsoft Power BI

Prof. Neelam Joshi[1], Pravin A Bhosale[2], Amod S Deshmukh[3], Samarth C Jedage[4], Anas A. H Shaikh[5]
Assistant Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2,3,4,5]
Sinhgad Institute of Technology, Lonavala, Maharashtra, India

Abstract: *Nowadays, a huge amount of crime and fraud takes place. Similarly, fraud in credit cards occurs when a transaction is being done. To evaluate and visualize credit card fraud transactions we implement this project. Detecting Anomaly (Outliers) in credit card transactions using Pycaret library. Microsoft Power BI is the most trending tool to prepare dashboards and reports for getting meaningful insight from the data and helps to take business-driven decisions. Hence Microsoft helps to visualize anomalies in credit card transactions. Anomaly Detection is one of the important and new features of Microsoft Power BI. Pycaret being an open-source low-code machine library in python helps to detect the anomaly. The output which contains the outliers will be represented in the form of illustrative visuals, which will be easy to understand and interpret. Since the number of transactions is done throughout the year. Developing a Dashboard that helps to visualize the total history of transactions based on the dataset. Creating reports based on credit card transactions for a time span. With amazing visuals and creativity, it becomes quick to understand what the data is telling the stakeholder or company management*

Keywords: Credit Card, Anomaly, Pycaret, Power BI.

I. INTRODUCTION

A credit card is a thin handy plastic card that contains identification information analogous as a hand or picture, and authorizes the person named on it to charge purchases or services to his account- charges for which he will be billed periodically. At moment, the information on the card is read by automated teller machines(ATMs), store reader, bank and is also used in online internet banking system. They have a unique card number which is of utmost significance. Its security relies on the physical security of the plastic card as well as the privacy of the credit card number. There is a rapid-fire- fire growth in the number of credit card deals which has led to a substantial rise in fraudulent exertion.

Credit card fraud is a wide- ranging term for theft and fraud committed using a credit card as a fraudulent source of finances in each trade. Generally, the statistical styles and multitudinous data mining algorithms are used to break this fraud discovery problem. Utmost of the credit card fraud discovery systems are predicated on artificial intelligence, Meta knowledge and pattern matching. The heritable algorithms are evolutionary algorithms which aim to gain the better results in barring the fraud. A high significance is given to develop effective and secure electronic payment system to descry whether a trade is fraudulent or not.

II. LITERATURE REVIEW

It is through diverse research on anomaly detection or outliers that one can gain a better understanding of the fraud situation that is faced by credit card companies. Pycaret's low-code nature made it suitable for machine learning models that needed to be trained quickly without wasting time in coding. We live in a data-driven world, and Microsoft Power BI gives us the ability to visualize and present data in a very modern way. Power BI from Microsoft is a trending visualization tool aimed at helping stakeholders to make more money. A combination of the Microsoft tool and anomaly detection coupled with pycaret is used to show the fraud in credit card transactions in this paper.

In this paper, we will concentrate on credit card fraud and credit card detection. A credit card fraud occurs when one individual uses other individual card for their particular use without the knowledge of its proprietor.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

III. ANOMALY DETECTION

Anomaly Detection is a technique in machine learning used for identifying rare items, events or observations which raise suspicions by differing significantly from the majority of the data.

Typically, the anomalous items will translate to some kind of problem such as bank fraud, a structural defect, medical problems or error.

There are three ways to implement an anomaly detector:

A. Supervised: Used when the data set has labels identifying which transactions are anomaly and which are normal. (This is similar to a supervised classification problem).

B. Semi-Supervised: The idea behind semi-supervised anomaly detection is to train a model on normal data only (without any anomalies). When the trained model is then used on unseen data points, it can predict whether the new data point is normal or not (based on the distribution of the data in the trained model).

C. Unsupervised: Exactly as it sounds, unsupervised means no labels and therefore no training and test data set. In unsupervised learning a model is trained on the complete dataset and assumes that most of the instances are normal. While looking for instances that seem to fit least to the remainder. There are several unsupervised anomaly detection algorithms such as Isolation Forest or One-Class Support Vector Machine. Each has their own method of identifying anomalies in the dataset

IV. PYCARET

PyCaret is an open-source, low-code machine learning library in Python that aims to reduce the hypothesis to insights cycle time in an ML experiment. In comparison with the other open-source machine learning libraries, PyCaret is an alternate low-code library that can be used to perform complex machine learning tasks with only a few lines of code.

PyCaret is simple and easy to use.PyCaret and its Machine Learning capabilities are seamlessly integrated with environments supporting Python such as Microsoft Power BI, Tableau, Alteryx, and KNIME to name a few.

V. MICROSOFT POWER BI

Microsoft Power BI is a business intelligence platform that provides nontechnical business users with tools for aggregating, analyzing, visualizing, and sharing data. Power BI's user interface is simple for users familiar with Excel and its deep integration with other Microsoft products makes it a very versatile self-service tool.

Microsoft Power BI is used to find insights within an organization's data. Power BI can help connect disparate data sets, transform, and clean the data into a data model and create charts or graphs to provide visuals of the data. All of this can be shared with other Power BI users within the organization.

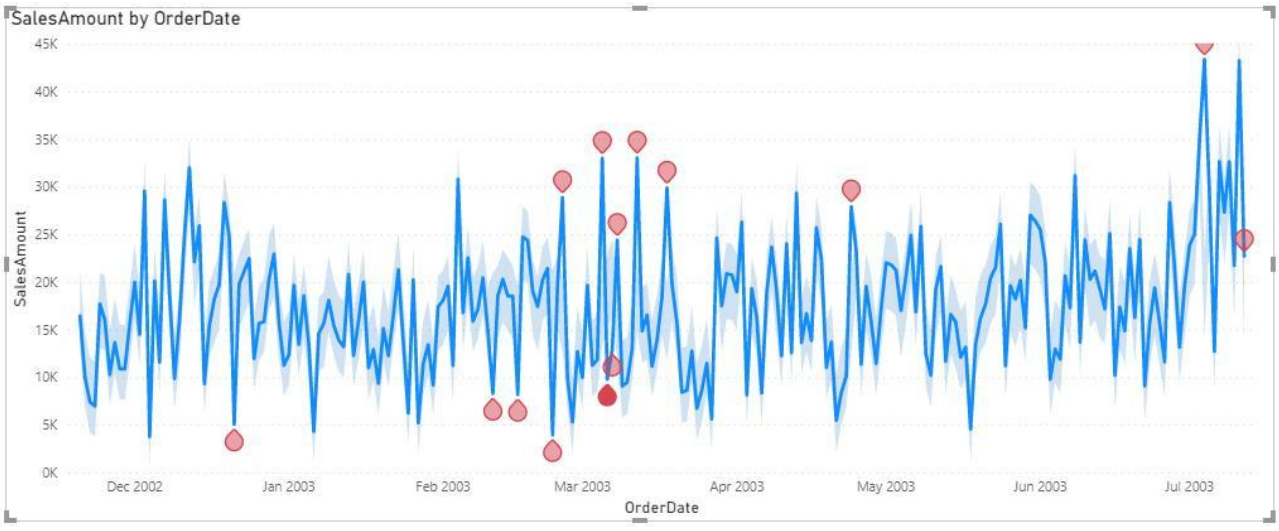

Fig.1. Line Chart of Anomaly Detection in Power BI

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

VI. WORKFLOW

Step 1 — Create an anaconda environment.
To create a virtual environment, we can use anaconda prompt in our laptop.

* Open anaconda prompt and type the below code in the CLI.
  conda create --n yourenvname python=(version)
* After pressing enter on the command, a little later you will get a prompt on the CLI.
* Type 'y' and press enter againand your environment will be created
* To activate the environment.
  conda activate yourenvname
* Notice after inputting the command our created environment was activated

Step 2 — Install PyCaret
Execute the following code in Anaconda Prompt:
* pip install pycaret (Installation may take 15–20 minutes)

Step 3 — Set Python Directory in Power BI
Setting the directory in power BI can be done using Global Settings in Power BI Desktop
* (File → Options → Global → Python scripting)
Set the path here Anaconda is installed.

Step 4 — Get Data
Getting the data from the excel/web/etc

Step 5 — Clean the Data
Cleaning the data in Power BI in Power Query Editor

Step 6 — Train the Model
Training the Model in Power Query Editor using python library in Run Python Script tab

Step 7 — Dashboard
Creating Dashboard and report to visualize the anomaly in the data.

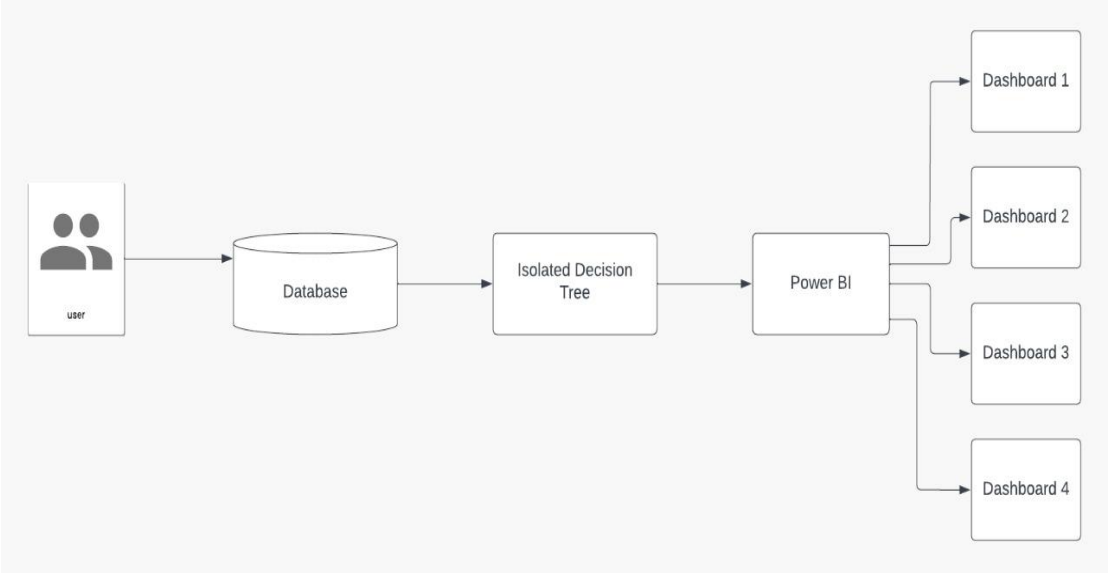

Fig. 2. Flowchart

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

VII. PROJECT IMPLEMENTATION

Anomaly detection in credit card transactions is the process of identifying unusual or suspicious activity in credit card transactions. This is important for detecting and preventing fraudulent transactions, which can be costly for both the cardholder and the financial institution.

Microsoft Power BI is a business analytics service that provides interactive visualizations and business intelligence capabilities with an interface that is user-friendly. Power BI can be used to analyse credit card transactions data and detect anomalies.

To perform anomaly detection in credit card transactions using Power BI, the first step is to obtain the transaction data from the credit card company. This data can then be imported into Power BI for analysis. Power BI provides a wide range of visualization tools that can be used to analyse the data, including bar charts, line charts, scatter plots, and histograms.

To detect anomalies, various statistical techniques can be applied to the data. These include clustering, regression, and machine learning algorithms such as decision trees and neural networks. Power BI provides built-in machine learning models that can be used for anomaly detection, as well as the ability to import custom models created using other tools such as Python or R.

Once anomalies have been detected, Power BI can be used to create alerts and notifications that can be sent to the appropriate parties for further investigation. Additionally, Power BI can be used to create reports and dashboards that provide a visual representation of the anomalies and trends in the data.

In summary, using Microsoft Power BI for anomaly detection in credit card transactions can help financial institutions to detect and prevent fraudulent activity, saving money and protecting their customers. Power BI provides a user-friendly interface, a wide range of visualization tools, and the ability to apply statistical and machine learning algorithms to the data for accurate detection of anomalies.

VIII. TOOLS AND TECHNOLOGIES USED

Microsoft Power BI Desktop

Microsoft Power BI is a powerful business analytics service that allows users to visualize and analyse data from various sources in a user-friendly interface. With Power BI, users can create interactive reports and dashboards, collaborate with colleagues, and share insights across their organization. Power BI provides a wide range of data visualization tools, including charts, graphs, maps, and tables, as well as advanced analytics capabilities such as machine learning and natural language processing. It also integrates seamlessly with other Microsoft products, such as Excel and Azure, making it easy to use and customize to meet the needs of any organization. Overall, Power BI is a comprehensive and versatile solution for data analysis and business intelligence.

"PyCaret" library in python

PyCaret is a low-code machine learning library in Python that simplifies the end-to-end machine learning process. It provides an easy-to-use interface for building and deploying machine learning models, data pre-processing, model selection, and hyperparameter tuning. PyCaret supports a wide range of machine learning algorithms and is designed to be beginner-friendly while also allowing for customization by experienced data scientists.

Anaconda distribution

Anaconda Distribution is a popular open-source data science platform that includes a distribution of Python, the conda package manager, and a collection of pre-installed data science libraries and tools. It is designed to make it easy to install and manage multiple data science packages and environments and is widely used by data scientists and analysts.

Microsoft Power BI Service

Power BI Service is a cloud-based business analytics service provided by Microsoft that allows users to create, view, and share interactive dashboards and reports. It provides a web-based interface that allows users to access their data and visualizations from anywhere and collaborate with others in real-time.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

IX. ALGORITHM DETAILS

Isolation Forest

Isolation Forest (iForest) is a machine learning algorithm for anomaly detection that uses a tree-based approach to identify outliers in a dataset. It is based on the principle that anomalies are points in a dataset that are isolated from the majority of the data points.

The iForest algorithm works by constructing a collection of random decision trees on a given dataset. Each tree is built by randomly selecting a feature and a split point to partition the data. The height of the tree corresponds to the number of splits required to isolate a point from the rest of the data.

To identify anomalies, the iForest algorithm calculates an anomaly score for each data point based on the average path length of the point in all the trees in the forest. The intuition is that an outlier will have a shorter average path length, as it can be isolated from most of the data in fewer splits.

One of the advantages of the iForest algorithm is its ability to handle high-dimensional data, as it does not rely on distance measures or density estimation. It also has a relatively low computational cost, making it suitable for large datasets. The iForest algorithm has been successfully applied in various domains, including cybersecurity, finance, and industrial systems.

X. CONCLUSION

Credit card fraud is the most common problem resulting in the loss of a lot of money for people and loss for some banks and credit card companies. This project wants to help people from their wealth loss and for the banked company and try to develop the model which more efficiently separates the fraud. Microsoft Power BI does a phenomenal job to detect anomalies in credit card transactions. Also, to visualize the credit card transaction that helps to take customer-driven decisions for the bank.

There are many questions within unsupervised anomaly detection and ML prediction that needs further studies. First, concerning unsupervised learning, the Isolation Forest, it would be of interest to see how the model originating from the training set will perform when injecting new anomalous instances from a different distribution into the test set. Second, oversampling and under sampling methods could be considered and see how the unsupervised Isolation would perform during those circumstances. Third, the anomaly scores from the Isolation Forest can be further examined and see how well these describe anomalous observations at the border of the decision region and also consider other kernel functions.

Microsoft Power BI is a trending tool nowadays, further, the data will gather in huge amounts. To sort and evaluate the data BI tools have a brilliant future

REFERENCES

[1]. Dr. Urmila R. Pol 1 And Dr. Tejshree U. Sawant 2, Automl: Building An Classification Model With Pycaret, YMER || ISSN : 0044-0477

[2]. Meenu, Swati Gupta, Sanjay Patel, Surender Kumar, Goldi Chauhan International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume-8, Issue-3, May 2020

[3]. S P Maniraj, Aditya Saini, Swarna Deep Sarkar Shadab Ahmed Credit Card Fraud Detection using Machine Learning and Data Science Assistant Professor (O.G.) International Journal of Engineering Research & Technology (IJERT). Published by : www.ijert.org Vol. 8 Issue 09, September-2019

[4]. Vijay Krishnan S, Bharanidharan G, Krishnamoorthy Research Data Analysis with Power BI 11th International CALIBER-2017 Anna University, Chennai, Tamil Nadu 02-04 August 2017 © INFLIBNET Centre, Gandhinagar, Gujarat

[5]. Asheesh Kumar Dwivedi, Ashish Kumar Rai, Ashish Kashyap Fraud Detection in Credit Card Transactions using Anomaly Detection Turkish Journal of Computer and Mathematics Education Vol.12 No.12 (2021), 837-846

[6]. Jain, Y. & Tiwari, N. & Dubey, S. & Jain, Sarika. (2019). "A comparative analysis of various credit card Fraud detection techniques" in International Journal of Recent Technology and Engineering. 7. 402-407.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

[7]. Isolation forests for anomaly detection improve fraud detection.", Blog Total Fraud Protection, 2019 [Online]. 18. https://blog.easysol.net/using-isolation-forests-anamoly-detection/ [Accessed 06 December 2020].

[8]. Pourhabibi, T.; Ong, K.-L.; Kam, B.H.; Boo, Y.L. Fraud detection: A systematic literature review of graph-based anomaly detection, approaches. Decis. Support Syst. 2020, 133, 113303. [CrossRef]

[9]. Changjun Jiang, et al. "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism." IEEE Internet of Things Journal, 5 (2018), pp. 3637-3647

[10]. Credit card fraud detection using Machine Learning Techniques John O. Awoyemi, Adebayo O. Adetunmbi, Samuel A. Oluwadare IEEE 2017

[11]. https://www.sumproduct.com/blog/article/power-bi-tips/power-bi-blog-anomaly-detection