IBM SPSS Modeler

Project Report

On

"Analysis of Fraud Prediction"

Submitted by: Abhishek Kumar

Raja Sardar

Ram Gope

Course: B. Tech CSE (AI & DS)

Semester: 5th

Institution: Arka Jain University

Guide: Mr. Prajwal Deep Khokhar

Date: 25-10-2025

Index

Project Brief

This project aims to detect **credit card fraud** using IBM SPSS Modeler in a simple, clear, and practical way. Fraud in online transactions happens very quickly, so identifying suspicious patterns is important for banks and customers.

In this project, we use a dataset containing credit card transactions and apply different machine-learning models to find which model can best separate **fraud** and **non-fraud** transactions. SPSS Modeler makes this process easier by allowing us to build workflows visually without coding.

The project includes:

- Loading the dataset (creditcard.csv).
- Doing EDA (Exploratory Data Analysis) to understand the data.
- Splitting the data into training and testing sets.
- Training multiple ML models like Decision Tree, Random Forest, Logistic Regression, Neural Network, and Anomaly Detection.
- Testing and comparing the performance of these models.

The main goal is to create a clear fraud detection model that can help identify unusual or risky transactions quickly. This project also helps in understanding how machine-learning workflows function inside IBM SPSS Modeler

Introduction

Credit card fraud is one of the biggest challenges in today's digital world. As online payments increase, the chances of fraudulent activities also rise. Detecting fraud early helps banks save money and protects customers from financial loss.

Machine learning plays a very important role in identifying unusual patterns in transaction data. Instead of manually checking each transaction, ML models can quickly analyze thousands of records and highlight suspicious ones.

In this project, IBM SPSS Modeler is used because it provides an easy, visual, and beginner-friendly workflow. It allows us to drag and drop nodes for data cleaning, model building, and testing—making the entire fraud detection process simple to understand.

This introduction gives a clear idea of what the project is about: using data and machine learning to detect fraud in a smooth and practical way.

Feasibility Study

Before starting any project, it is important to check whether the project is practical, affordable, and easy to implement. This is called a **Feasibility Study**. For this credit card fraud detection project, we look at three types of feasibility: technical, economic, and operational.

a. Technical Feasibility

- IBM SPSS Modeler is a powerful tool that supports different machine-learning models.

- It allows easy data loading, data cleaning, visualization, and model building using simple drag-and-drop nodes.

- No programming knowledge is required, making it suitable for beginners.

- The laptop/computer requirements are basic, and the dataset used is not heavy, so performance is smooth.

b. Economic Feasibility

- The dataset (creditcard.csv) is **free and publicly available**, so there is no cost for data.

- SPSS Modeler is available under academic or student licenses, making it affordable for learning projects.

- Since no paid cloud resources or advanced hardware are needed, the overall project cost is very low.

c. Operational Feasibility

- The workflow is simple and easy to understand for students and beginners.

- SPSS Modeler's visual interface helps in quickly building and modifying models.

- The system can be used to easily predict fraud on new transaction data.

- The project can be presented clearly with diagrams, making it suitable for academic submissions.

Overall, the project is **highly feasible**, cost-effective, and easy to execute using IBM SPSS Modeler.

Project Details

4.1 Dataset Description
- The dataset contains **31 fields**, including transaction amount, anonymized features (V1–V28), and a fraud label.
- "Class" field: 0 = normal transaction, 1 = fraud.

4.2 EDA (Exploratory Data Analysis)
- Obtained summary statistics using the **Statistics Node**.
- Checked distributions, correlations, and missing values.
- Visualized fraud vs non-fraud ratio (highly imbalanced dataset).

4.3 Training Models Workflow
The training area includes:
- **Data Partition** (60% for training).
- Multiple models connected to the training data, such as:
  - Decision Tree
  - Random Forest
  - Logistic Regression
  - Neural Network
  - Anomaly Detection
- Each model produces a T/F output (fraud vs non-fraud).
- Performance is evaluated using accuracy and ROC.

4.4 Testing Models
- Remaining 40% data is used for testing.
- Each trained model is applied on the test dataset.
- Analysis nodes provide performance comparison.

Conclusion / Summary

This project successfully demonstrates a credit card fraud detection system using IBM SPSS Modeler. Multiple models were trained and tested. The comparison helps identify the most accurate model for fraud prediction. The visual workflow makes the project easy to understand, implement, and present.

Overall Outcome: Machine-learning models provide an efficient solution for detecting fraudulent transactions and improving financial security.