DEPARTMENT OF ENGINEERING

Predictive Analysis Project File

Title: Student Academic Performance Prediction using SPSS Modeler

Submitted by: Prince Kumar

Enrollment No.: AJU/231816

Roll No.: BTCS/DS-042

Semester: 5th

Index:

Serial No.	Content
1.	Introduction
2.	Project Overview
3.	Feasibility Study
4.	Project Details
5.	Conclusion Summary

INTRODUCTION:

This project focuses on developing a Student Academic Performance Prediction system using IBM SPSS Modeler, a powerful data mining and predictive analytics tool. SPSS Modeler enables seamless data preprocessing, exploration, feature selection, and application of machine learning techniques such as decision trees, neural networks, and regression models. By training predictive models on historical student data, the system estimates future academic performance and highlights influencing factors.

The primary goal of this project is to support educators, administrators, and academic planners in making proactive interventions, designing personalized learning strategies, and enhancing overall academic quality. Ultimately, this predictive framework contributes to early identification of performance issues, improved student retention, and a more efficient educational ecosystem.

PROJECT OVERVIEW:

This project focuses on predicting student academic performance using predictive analytics techniques in IBM SPSS Modeler. The dataset used contains information about 395 students and includes 33 attributes related to demographics, family background, social behaviour, lifestyle, and academic history. Key variables such as age, gender, study time, parental education, school support, extracurricular activities, alcohol consumption, absences, and previous grades (G1, G2) serve as potential predictors. The final academic outcome—G3 (final grade)—is used as the target variable for prediction. The project follows a complete predictive analysis workflow, including data preprocessing, missing value handling, variable transformation, and exploratory statistical analysis. SPSS Modeler is used to build and compare multiple predictive models such as Decision Trees, Linear Regression, Neural Networks, and Logistic Regression to identify the most accurate model. Evaluation metrics like accuracy, RMSE, ROC curves, and lift charts help measure model performance and reliability.

FEASIBILITY STUDY:

1. **Technical Feasibility**

- The project requires IBM SPSS Modeler, which supports data preprocessing, statistical modeling, and predictive analytics.
- The dataset contains relevant academic, demographic, behavioural, and past performance variables, making it suitable for machine learning prediction.
- Hardware requirements are minimal—standard system with 8GB RAM, i5 processor is sufficient.
- No advanced programming knowledge is mandatory because SPSS Modeler provides a drag-and-drop visual interface.

2. Operational Feasibility

- Educational institutions, teachers, and administrators can easily interpret prediction insights and use them for academic planning.
- The system can help identify at-risk students early, improving retention and academic outcomes.
- Stakeholders can integrate results into existing academic workflows (counseling, mentoring, remedial classes).

3. Economic Feasibility

- SPSS Modeler may require licensing, but student/educational licenses may reduce cost.
- Dataset is freely available—no data acquisition expense.
- Development cost is low since it requires only analytical effort and basic computing resources.
- The value gained—better student performance, reduced dropout rates—outweighs the cost.

6. Organizational Feasibility

- Institutions already maintain student academic data, making model integration easy.
- Staff can be trained quickly to interpret prediction dashboards and reports.
- Decision-makers gain actionable insights for academic strategies.

PROJECT DETAILS:

1. Project Title

Student Academic Performance Prediction Using Predictive Analytics in IBM SPSS Modeler.

2. Project Domain

Educational Data Mining, Predictive Analytics, Machine Learning

3. Project Purpose

To develop a prediction model that estimates students' final academic performance and identifies key factors influencing success or risk, enabling educational institutions to take proactive measures.

4. Dataset Description

- **Total Records:** 395 students
- **Attributes:** 33 variables including
    - **Demographic factors:** age, gender, address
    - **Family background:** parental education, family size, relationship status
    - **Behavioural patterns:** alcohol consumption, social activities, absences

  - **Academic indicators:** study time, failures, school support, past grades (G1, G2)
- **Target Variable: G3 – Final Grade**
- Dataset is well-structured, numeric/categorical, and suitable for classification or regression analysis.

5. Tools & Technologies Used

- IBM SPSS Modeler — Data preparation, modeling, visualization
- Microsoft Excel/Google Sheets — Data review (optional)
- Statistical & predictive modeling techniques:
  - Decision Trees (C&R Tree, CHAID)
  - Linear/Multiple Regression
  - Neural Networks
  - Logistic Regression
  - Random Forest (if applied)

9. Advantages

- Data-driven decision making
- Supports proactive educational intervention
- Reduces academic failure rate
- Easy model deployment due to SPSS interface

10. Limitations

- Dataset size may limit generalization
- Performance may vary across institutions
- Social/psychological factors may not be fully captured

CONCLUSION:

The Student Academic Performance Prediction project successfully demonstrates how predictive analytics can be applied to educational data to forecast student outcomes. Using the provided dataset and IBM SPSS Modeler, multiple predictive models were developed, analyzed, and evaluated to determine the most accurate approach for estimating final grades

(G3). The results show that factors such as previous academic scores (G1, G2), study time, family support, and student absenteeism significantly influence academic performance.

Here's my project :