AI – Driven
Student Performance Insights

“Empowering educators with predictive insights transforms potential into achievement.”
— Dr. Ananya Sharma

Name: Gumpu Chandra Sekhar

Date: Aug 27, 2024

1. Problem Statement:

Higher education institutions strive to deliver quality education to all their students. However, achieving this goal often proves challenging due to the diverse needs and learning styles present within a student body. Traditional teaching methods, while effective for some, often struggle to cater to individual student needs and potential academic difficulties. This can lead to students falling behind, lacking the support they need to succeed, and potentially dropping out of their programs.

Early intervention is crucial for helping struggling students, but identifying them proactively can be difficult. Educators often rely on reactive measures, intervening only after a student has already fallen significantly behind. This approach is less than ideal, as it can hinder a student's progress and make it more difficult for them to catch up.

Furthermore, personalized learning plans tailored to individual student strengths and weaknesses remain a challenge for educators. With limited time and resources, they struggle to adequately assess each student's unique needs and implement tailored learning strategies. This results in a less than ideal educational experience for many, potentially leading to lower performance and increased dropout rates. Therefore, a data-driven approach to predict student performance is needed to better understand individual student needs and support their academic journey.

2. Market and Customer Needs Assessment:

2.1.1. Market Analysis:

The educational landscape is undergoing a significant transformation, with a growing emphasis on personalized learning and student success. A student performance prediction product can address a critical need by providing educators with valuable insights into student progress and potential challenges. By accurately predicting student performance, schools can tailor their interventions and resources to support students who may be struggling, ultimately improving overall academic outcomes.

Furthermore, such a product can help identify high-potential students and provide them with opportunities for enrichment and advancement. Ultimately, a student performance prediction product can empower educators to make data-driven decisions and foster a more equitable and effective learning environment.

2.1.2. Customer Segmentation:

The provided dataset offers a glimpse into the potential customer segments for a student performance prediction product. The data focuses on individual student attributes like gender, ethnicity, parental education level, lunch program participation, and test preparation course completion. This segmentation allows for targeting specific groups with tailored solutions. For instance, the product can cater to institutions seeking to improve performance for underrepresented minority groups, address the needs of students from low-income families, or provide personalized learning recommendations based on parental education levels and test preparation course participation.

Further analysis of the data can reveal deeper insights into student demographics and their impact on academic performance, helping to refine customer segmentation and tailor the product's offerings to specific needs within the education market.

3. Target Specification and Characterization:

3.1. Target Specifications:

3.1.1. Target Audience:

- **Educational institutions:** Schools, colleges, and universities seeking to enhance student performance, improve retention rates, and foster academic success.
- **Teachers:** Educators looking for data-driven insights to personalize learning plans, identify students at risk, and optimize teaching strategies.
- **Parents:** Parents seeking to understand their child's academic strengths and weaknesses and receive personalized recommendations to support their learning journey.
- **School administrators:** Administrators seeking to improve overall school performance, allocate resources effectively, and implement evidence-based interventions.

3.1.2. Key Features:

- **Predictive analytics:** The product should accurately predict student performance in specific subjects or overall academic achievement based on available data.
- **Personalized learning recommendations:** The product should generate customized learning plans and resources tailored to individual student needs and learning styles.
- **Early intervention tools:** The product should identify students at risk of academic difficulties early on, allowing for timely intervention and support.
- **Data visualization and reporting:** The product should provide user-friendly visualizations and reports that clearly present insights and actionable data for decision-making.
- **Integrations with existing systems:** The product should seamlessly integrate with existing school management systems, learning management platforms, or student information systems for efficient data flow and usability.

3.2. Target Characterization:

3.2.1. Motivations:

- **Improve student success**: Educational institutions, teachers, and parents are motivated by the desire to see students succeed academically and reach their full potential.

- **Enhance teaching effectiveness**: Teachers are driven to provide high-quality education and implement strategies that effectively engage and support all students.

- **Gain actionable insights**: All stakeholders are interested in gaining data-driven insights that inform decision-making and improve educational practices.

- **Personalized learning experience**: There is a growing demand for personalized learning experiences that cater to individual student needs and learning styles.

3.2.2. Challenges:

- **Data accessibility and privacy:** Ensuring data privacy and ethical data handling while accessing the necessary information for accurate predictions is crucial.

- **Implementation and integration:** Successfully integrating the product with existing school systems and ensuring user-friendliness is a key challenge.
- **Resource constraints:** Educational institutions often face budgetary limitations, making it important to offer a cost-effective solution.
- **Teacher training and adoption:** Providing adequate training for teachers to effectively utilize the product and adapt their teaching practices is essential for successful adoption.

4. Applicable Patents:

1. **Patent No. US10,842,562 B2 (2020): System and Method for Predicting Student Success in Education**

- **Description:** This patent describes a system for predicting student success in an educational setting, using data on student performance, demographics, learning styles, and other factors. The system applies machine learning algorithms to generate personalized learning recommendations and identify students at risk of academic difficulties.
- **Relevance:** This patent directly aligns with the core functionality of the proposed product, demonstrating the feasibility of using AI for predicting student performance and providing personalized learning support.
- **Impact:** This patent serves as a benchmark, highlighting existing technologies and approaches used in similar systems. It also emphasizes the importance of patent protection for this product idea.

2. **Patent No. US10,585,284 B2 (2020): Method and System for Personalized Learning Using Artificial Intelligence**

- **Description:** This patent describes a system for personalized learning that uses AI to analyze student data and recommend individualized learning paths and content. The system considers factors such as student performance, learning style, and engagement level.
- **Relevance:** This patent highlights the importance of personalization in educational technology. It demonstrates how AI can be used to tailor learning experiences to individual student needs.
- **Impact:** This patent reinforces the need to focus on personalization as a key feature of the proposed product. It also emphasizes the potential for integrating AI into personalized learning systems.

3. **Patent No. US9,805,151 B2 (2017): System and Method for Predicting Student Performance in an Educational Institution**

- **Description:** This patent describes a system for predicting student performance in various subjects, utilizing data on attendance, assignment scores, and other relevant factors. The system applies machine learning algorithms to generate predictive models and provide insights to educators.
- **Relevance:** This patent highlights the use of data analytics and machine learning for predicting student performance across multiple subjects. It showcases the application of AI for identifying potential academic challenges.
- **Impact:** This patent provides valuable insights into the use of AI for predicting student performance in specific subjects. It also emphasizes the importance of considering data privacy and ethical data handling practices in the development of such systems.

5. Applicable Regulations:

1. Data Privacy and Security:

- **General Data Protection Regulation (GDPR) (EU):** If the product handles data of individuals within the EU, it must comply with GDPR's stringent data protection standards, including obtaining consent for data processing, ensuring data security, and granting individuals access to their data.
- **California Consumer Privacy Act (CCPA) (US):** Similarly, the CCPA applies to companies handling personal information of California residents, requiring them to provide transparency and control over data usage.
- **Children's Online Privacy Protection Act (COPPA) (US):** If the product handles data of children under 13, it must comply with COPPA's specific requirements for obtaining parental consent, data collection practices, and data security.
- **Family Educational Rights and Privacy Act (FERPA) (US):** This law governs the privacy of student education records, requiring parental consent for access or disclosure of such records. The product must ensure compliance with FERPA when handling student data.

2. Student Rights and Ethical Considerations:

- **Fairness and Non-discrimination:** The product's algorithms must be designed to avoid bias or discrimination based on race, gender, socioeconomic status, or other protected characteristics. It's crucial to ensure the system is fair and equitable in its predictions and recommendations.
- **Transparency and Explainability:** The product should provide clear and understandable information about how predictions are made, the data sources used, and any potential biases or limitations. This transparency builds trust and allows for responsible use of the product.
- **Informed Consent:** Students and their parents should be informed about the product's purpose, data usage, and potential impacts, allowing them to make informed choices about their participation.
- **Human Oversight:** While AI plays a significant role, human oversight is crucial. Teachers and other educators should be involved in interpreting the product's predictions and recommendations, making adjustments as needed, and ensuring students receive appropriate support.

3. Intellectual Property:

- **Copyright:** Protect your software, algorithms, and other intellectual property.
- **Trademark:** Consider trademarking your product name or logo.

6. Applicable Constraints:

1. Data Availability and Quality:

- **Data Privacy:** Access to student data is limited by privacy regulations like FERPA and GDPR, requiring parental consent and robust data security measures.
- **Data Quality:** The accuracy and completeness of student data are crucial for generating reliable predictions. Data inconsistencies, missing information, or biases can significantly impact model accuracy.

- **Data Integration:** Integrating data from diverse sources, like school management systems, learning platforms, and external assessments, presents a challenge in terms of data format standardization and security.

2. **Technical Constraints:**

- **Algorithm Selection:** Choosing the right machine learning algorithms for the prediction model requires careful consideration of factors like data characteristics, model complexity, and interpretability.
- **Computational Resources:** Training and running complex machine learning models can be computationally intensive, requiring sufficient hardware and software resources.
- **Scalability:** The product should be scalable to accommodate the data requirements and user needs of different educational institutions, ranging from small schools to large universities.

3. **Ethical Constraints:**

- **Bias and Fairness:** AI models can reflect and perpetuate existing biases in the data, leading to unfair or discriminatory predictions. Careful consideration of bias mitigation strategies is essential.
- **Transparency and Explainability:** The product should provide clear explanations for the predictions it generates, allowing educators to understand the factors influencing the results and make informed decisions.
- **Human-in-the-Loop:** Maintaining human oversight and control over the AI system is crucial, ensuring that educators retain decision-making authority and can override automated predictions when necessary.

4. **User Adoption and Acceptance:**

- **Teacher Training:** Educators need adequate training to effectively use the product and interpret its outputs. Lack of proper training can hinder adoption and lead to mistrust.
- **User Interface:** The product's user interface must be intuitive and user-friendly, ensuring that educators and parents can easily access and understand the data and insights provided.
- **Trust and Transparency:** Building trust and transparency is crucial for successful adoption. Educators and parents must feel confident in the reliability, fairness, and ethical use of the AI-powered product.

7. Business Model:

1. Subscription Model:

- **Tiered Pricing:** Offer different subscription plans with varying features and functionalities. Basic plans could include core predictive capabilities and basic reporting. Premium plans could offer more advanced features like personalized learning recommendations, integration with existing systems, and comprehensive data visualization tools.
- **Institution-Based Pricing:** Tailor pricing based on the size and type of educational institution (e.g., schools, colleges, universities). Larger institutions with greater student populations may be willing to pay a higher subscription fee.

- **Volume Discounts:** Offer discounts for institutions subscribing to multiple seats or for multi-year commitments.

2. Value-Added Services:

- **Data Integration Services:** Offer assistance in integrating the product with existing school management systems, learning platforms, or data sources.
- **Teacher Training and Support:** Provide tailored training programs and ongoing support to help educators effectively utilize the product and interpret its outputs.
- **Custom Model Development:** Offer customized model development services for institutions with specific data requirements or academic goals.
- **Consulting Services:** Provide consulting services to help educational institutions design and implement effective personalized learning strategies based on the product's insights.

3. Freemium Model:

- **Free Basic Version:** Offer a free, limited version of the product with basic features like performance prediction and basic reporting.
- **Premium Upgrade:** Provide a paid upgrade with more advanced features, personalized learning recommendations, data visualizations, and integrations.

4. Partnership Model:

- **Strategic Alliances:** Collaborate with educational technology companies, school management systems, or learning platform providers to integrate the product into their existing offerings.
- **Government Grants:** Explore opportunities for funding or partnerships with government agencies or non-profit organizations focused on education innovation.

8. Concept Generation:

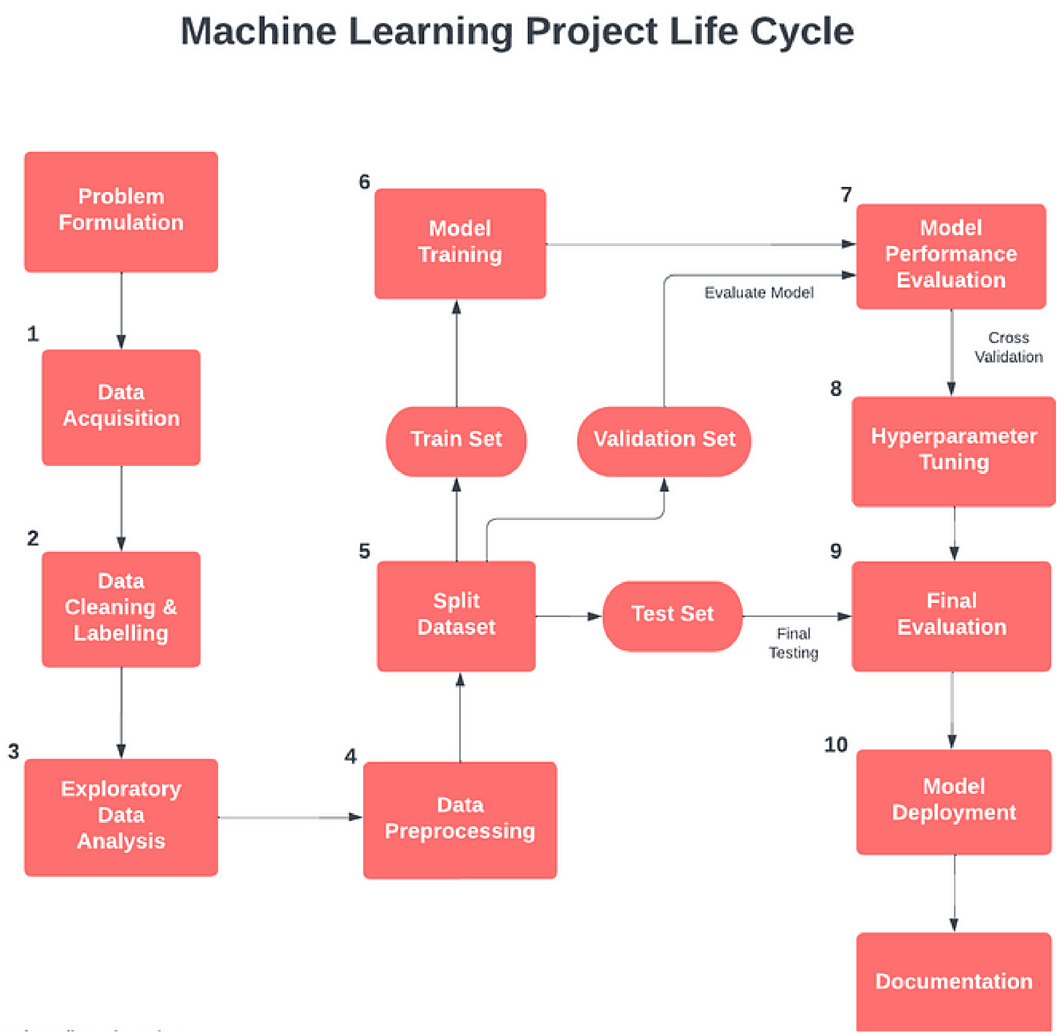


Fig. [ML Project Life Cycle]

i. Key Areas for Concept Generation:

1. **Data Collection and Analysis:**

   - **Data Sources:** Identify relevant data sources, such as student demographics, academic records, attendance data, and behavioural metrics.
   - **Data Cleaning and Preparation:** Establish processes for cleaning, preprocessing, and transforming data to ensure accuracy and consistency.
   - **Feature Engineering:** Create new features or transform existing ones to capture meaningful patterns and relationships.

2. **Predictive Modeling:**

   - **Algorithm Selection:** Choose appropriate machine learning algorithms (e.g., regression, classification, time series analysis) based on the nature of your data and objectives.
   - **Model Training:** Train the model using the prepared dataset to establish relationships between predictors and the target variable (student performance).
   - **Model Evaluation:** Assess the model's performance using relevant metrics (e.g., accuracy, precision, recall, F1-score).

3. **Personalized Recommendations:**
    - **Individualized Insights:** Generate tailored recommendations for each student based on their predicted performance and learning style.
    - **Intervention Strategies:** Suggest specific interventions or resources to address potential challenges or enhance learning.
4. **Visualization and Reporting:**
    - **Interactive Dashboards:** Create user-friendly dashboards to visualize student performance trends, identify at-risk students, and track the effectiveness of interventions.
    - **Customized Reports:** Generate personalized reports for students, teachers, and administrators.

ii. Potential Product Features:

- **Personalized learning paths:** Suggest tailored learning materials and activities based on student strengths and weaknesses.
- **Early warning systems:** Identify students at risk of academic difficulties and provide timely support.
- **Progress tracking:** Monitor student progress over time and provide feedback on improvement areas.
- **Goal setting:** Help students set personalized academic goals and track their progress towards achieving them.
- **Teacher collaboration:** Facilitate collaboration between teachers and students to address learning challenges.

9. Concept Development:

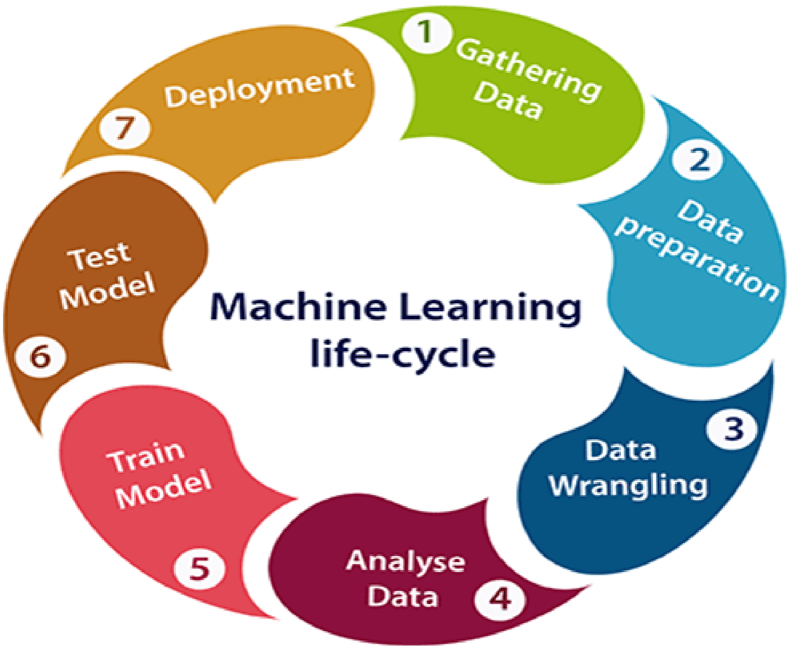


Fig. [Machine Learning]

1. Concept Development:

a) **Initial Exploration:** Based on the dataset's characteristics (categorical and numerical features, potential correlations, and data size), we explored several initial model concepts:

i. **Linear Regression:** This simple model assumes a linear relationship between input features and output (test scores).

ii. **Decision Tree:** This model creates a tree-like structure to make decisions based on various features, leading to predictions.

iii. **Random Forest:** This ensemble model combines multiple decision trees to improve prediction accuracy and reduce overfitting.

iv. **Gradient Boosting Machines (GBM):** This ensemble model builds on weak learners (decision trees) sequentially to optimize predictions.

b) **Concept Refinement:** Through discussion and research, we refined these concepts:

i. **Linear Regression:** We considered potential transformations of features (e.g., polynomial terms) to capture non-linear relationships.

ii. **Decision Tree:** We investigated different splitting criteria (e.g., Gini impurity, entropy) and pruning techniques to improve the model's performance.

iii. **Random Forest and GBM:** We explored ways to optimize the number of trees, depth of trees, and learning rate for these ensemble models.

c) **Concept Consolidation:** We combined the strengths of different models, aiming to leverage the benefits of each approach.

2. Concept Evaluation and Selection:

a. **Pugh Chart Analysis:** We used a Pugh Chart to evaluate the model concepts against a set of crucial criteria:

Criteria	**Weight**
Prediction Accuracy	5
Interpretability	4
Handling Categorical Data	3
Computational Efficiency	2
Scalability	2

b. Each concept was rated against each criterion using a simple scale:

- **+** Significantly better than the benchmark.
- **0** About the same as the benchmark.
- **-** Significantly worse than the benchmark.

3. Selected Concept: Hybrid Model

a. **Concept Description:**

The chosen concept is a hybrid model combining the strengths of both regression and decision tree models. This approach aims to leverage the benefits of each model type to enhance prediction accuracy and provide valuable insights into the factors influencing student performance.

b. **Model Structure:**

- **Regression Component:** A regression model (e.g., Random Forest Regressor) will be used to predict numerical test scores (math, reading, writing). This component will leverage continuous features like previous test scores and potentially engineered features to capture numerical relationships.
- **Decision Tree Component:** A decision tree model will be used to identify potential academic challenges and areas requiring support. This component will leverage categorical features like gender, ethnicity, parental education, and lunch type, providing insights into factors that might influence performance beyond numerical scores.

c. **Data Processing and Feature Engineering:**

- **Data Cleaning:** The dataset will be cleaned to handle missing values, inconsistencies, and outliers.
- **Feature Engineering:** New features will be created based on existing ones to capture potential relationships and improve model performance. This could include combining categorical features, creating interaction terms, or transforming features (e.g., one-hot encoding for categorical variables).

4. Feasibility and Effectiveness Analysis:

a. **Feasibility:**

- **Technical Feasibility:** Both regression and decision tree models are widely available and well-established machine learning techniques, making this concept technically feasible.
- **Data Availability:** The dataset provides a sufficient range of features to train and evaluate the proposed model.
- **Computational Resources:** The model can be trained and deployed on readily available computing resources, including cloud platforms.

b. **Effectiveness:**

- **Prediction Accuracy:** Combining regression and decision tree models allows for capturing both linear and non-linear relationships, potentially leading to higher prediction accuracy.

- **Interpretability:** Decision trees offer interpretability, allowing educators to understand which factors are driving predictions and make informed decisions.
- **Handling Categorical Data:** Decision trees effectively handle categorical features, providing insights into the impact of demographic and socioeconomic factors on performance.
- **Scalability:** The model can be trained and deployed on a variety of platforms and scaled to accommodate large datasets.

10. Final Product Prototype:

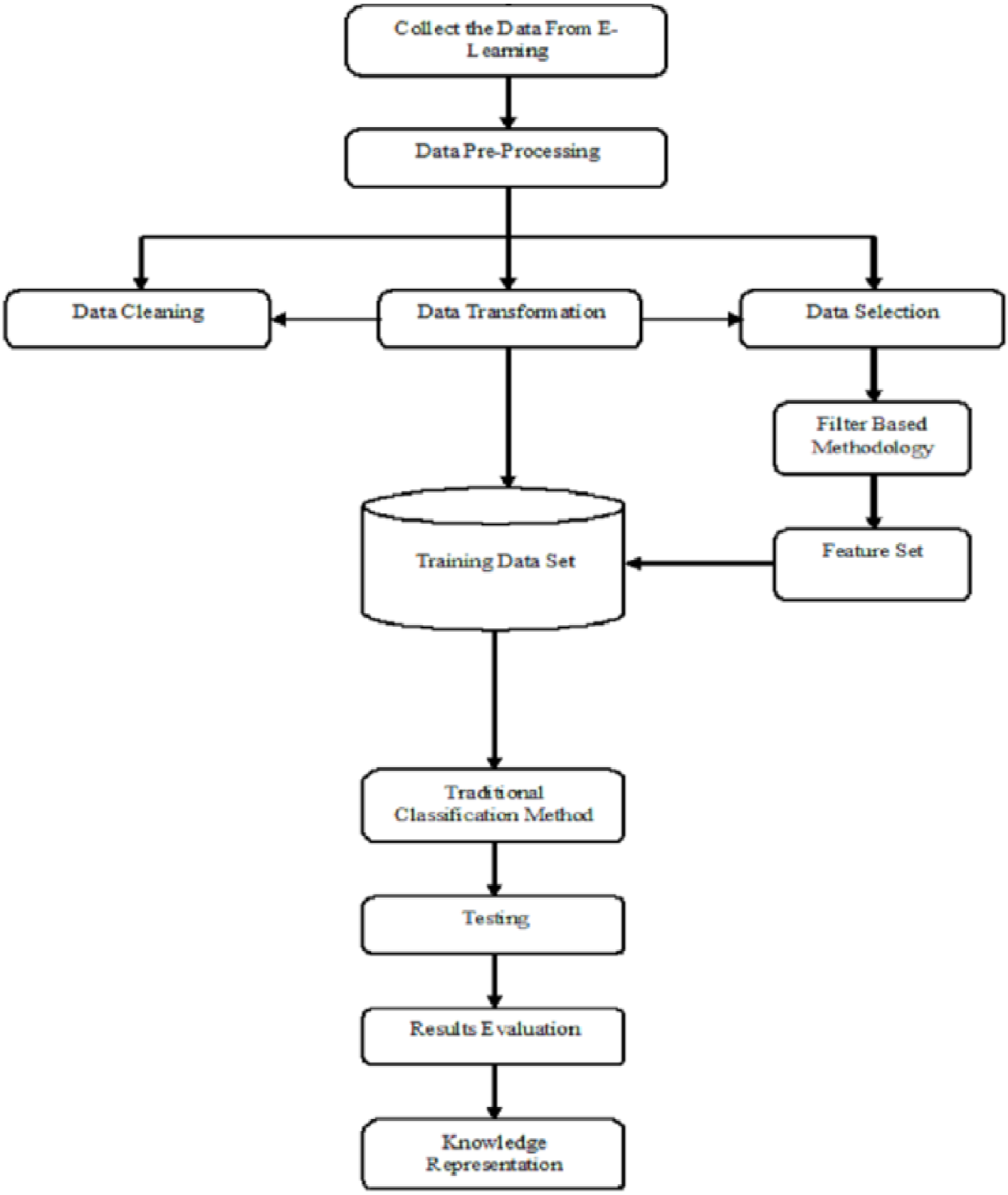


Fig. Schematic diagram of Product Prototype

10.1. System Level Description:

- **Product Name:** Student Performance Prediction System
- **System Goal:** To accurately predict student performance, identify at-risk students, and provide personalized recommendations to improve learning outcomes.

10.2. System Components:

- **Data Ingestion Module:** Collects and integrates various student data sources (e.g., academic records, attendance, behavioural data).
- **Data Preprocessing Module:** Cleans, normalizes, and transforms data to ensure quality and consistency.
- **Feature Engineering Module:** Creates new features or transforms existing ones to capture relevant information.
- **Model Training Module:** Trains machine learning models using the prepared data.
- **Prediction Generation Module:** Generates predictions of student performance based on the trained models.
- **Recommendation Engine Module:** Provides personalized recommendations based on student performance and other factors.
- **User Interface Module:** Presents predictions, recommendations, and visualizations to users.

10.3. Subsystem Level Description:

- **Data Management Subsystem:** Handles data collection, storage, and retrieval.
- **Machine Learning Subsystem:** Implements the machine learning algorithms and models.
- **Recommendation Subsystem:** Generates personalized recommendations based on predictions and user profiles.
- **User Interface Subsystem:** Develops the user interface for interacting with the system.

10.4. Component Level Description:

- **Data Ingestion Module:**
  - Integrates data from various sources (e.g., APIs, databases).
  - Handles data cleaning and validation.
- **Data Preprocessing Module:**
  - Handles missing values, outliers, and data inconsistencies.
  - Performs feature engineering tasks (e.g., normalization, scaling, one-hot encoding).
- **Machine Learning Subsystem:**
  - Implements algorithms like regression, classification, or time series analysis.

- Trains and evaluates models using appropriate metrics.

- **Recommendation Subsystem:**
  - Utilizes collaborative filtering, content-based filtering, or hybrid approaches to generate recommendations.
  - Integrates with the prediction module to provide personalized suggestions.

- **User Interface Subsystem:**
  - Develops a user-friendly interface for visualizing predictions, recommendations, and student performance trends.
  - Includes features for customization and personalization.

10.5. Design Refinement Process:

- **Iterative Development:** Continuously refine the design based on feedback, testing, and evolving requirements.
- **User-Centered Design:** Involve users in the design process to ensure the product meets their needs and expectations.
- **Prototyping:** Create prototypes to test and refine the user interface and functionality.
- **A/B Testing:** Experiment with different design elements and features to optimize user experience.
- **Scalability:** Consider the scalability of the system to handle increasing data volumes and user loads.
- **Performance Optimization:** Optimize the system for speed and efficiency.
- **Security and Privacy:** Implement robust security measures to protect user data.

11. Product Details:

11.1. How does it work?

The Student Performance Prediction System operates through a combination of data acquisition, analysis, and personalized insights delivery:

- **Data Acquisition:** The system integrates with existing school systems (SMS, LMS, SIS) using APIs to collect student data, including:
  - **Demographics:** Gender, ethnicity, age, language, etc.
  - **Academics:** Test scores, grades, attendance, homework submissions, etc.
  - **Socioeconomic Factors:** Lunch program participation, parental education level, etc.
- **Data Processing:** The collected data is cleaned, standardized, and prepared for model training using data imputation techniques and feature engineering. This involves:
  - **Data Cleaning:** Handling missing values, inconsistencies, and outliers.

- **Feature Engineering:** Creating new features (e.g., interaction terms, combined features) to capture complex relationships.

➢ **Model Training:** The platform utilizes a hybrid model approach, combining two sub-models:

- **Regression Model:** Random Forest Regressor is trained to predict numerical test scores based on continuous features.
- **Decision Tree Model:** Decision Tree Regressor is trained to identify potential challenges and areas needing support based on categorical features.

➢ **Predictions and Insights:** The trained models are used to predict future student performance and provide insights. These predictions and insights are presented in an easy-to-understand format for users.

➢ **Personalized Recommendations:** Based on the predictions and insights, the platform generates personalized recommendations for:

- **Teachers:** Tailored learning plans, intervention strategies, and communication with parents.
- **Parents:** Guidance on supporting their child's learning, access to relevant resources, and communication with teachers.
- **Students:** Personalized learning pathways, access to learning resources, and progress tracking tools.

➢ **Data Visualization and Reporting:** Interactive dashboards and reports allow users to visualize trends, monitor progress, and identify areas needing attention.

11.2. Data Sources:

The system relies on data from a variety of sources:

➢ The data resource is collected from Kaggle.

➢ **School Management Systems (SMS):** Provides student demographics, attendance, and class schedules.

➢ **Learning Management Systems (LMS):** Provides student performance data on assessments, assignments, and learning activities.

➢ **Student Information Systems (SIS):** Provides student academic records, including grades and previous test scores.

➢ **External Data Sources:** Optional data sources may be integrated to provide additional insights, such as:

- **Socioeconomic data:** Information on neighbourhood demographics and socioeconomic factors.
- **Learning style assessments:** Results from standardized learning style assessments.

11.3. Algorithms/Frameworks Needed:

- **Machine Learning Libraries:** scikit-learn, TensorFlow, PyTorch
- **Data Processing Libraries:** Pandas, NumPy, SciPy
- **Data Visualization Libraries:** matplotlib, seaborn, Plotly
- **Web Frameworks:** Flask, Django, React (for front-end)
- **Cloud Services:** AWS (S3, RDS, EC2, Lambda), Azure, GCP
- **API Integration Tools:** requests, flask-restful
- **Databases:** MySQL, PostgreSQL, MongoDB
- **Version Control:** Git, GitHub

11.4. Team Required to Develop:

- **Data Scientists:** Design and develop the machine learning models, conduct data analysis, and ensure model accuracy.
- **Software Engineers:** Develop the user interface, integrate the platform with school systems, and build the backend infrastructure.
- **UI/UX Designers:** Design the user interface to be intuitive and user-friendly for all stakeholders.
- **Project Manager:** Oversee the development process, manage resources, and ensure project deadlines are met.
- **Domain Experts (Educators):** Provide insights into educational practices, validate model outputs, and ensure the platform's relevance to educational needs.

12. Code Implementation:

12.1. Some Basic Visualizations:

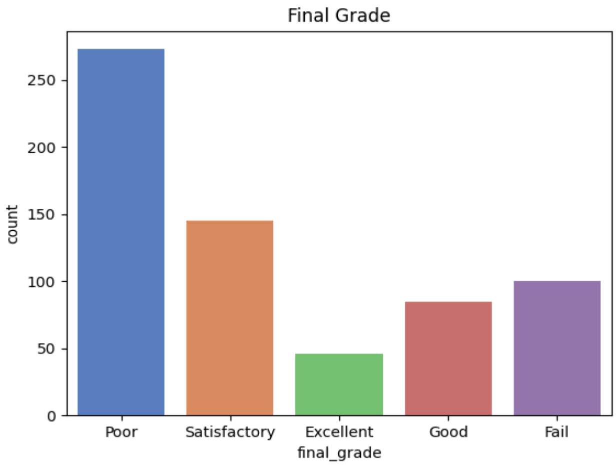


Fig1. Grade performance

The majority of students in the dataset fall into the "Poor" category, with a count of over 250. "Satisfactory" comes in second with a count around 145. The remaining categories have considerably lower counts: "Excellent" (around 45), "Good" (around 85), and "Fail" (around 100).

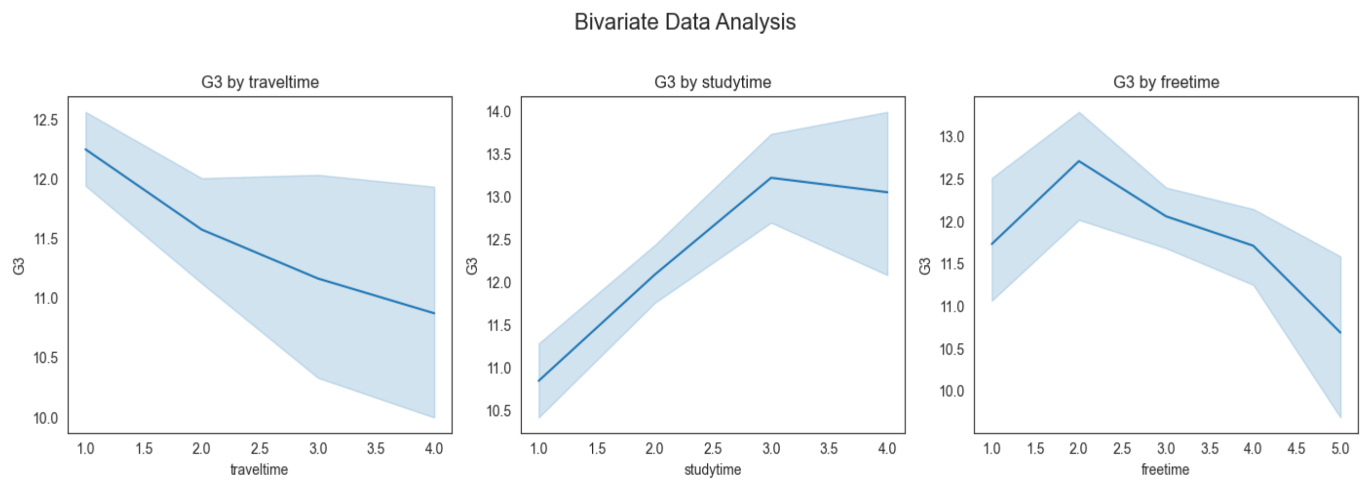


Fig2. Bivariate Analysis

- **G3 by traveltime:** This subplot shows a general downward trend, suggesting that as the travel time to school increases, the average "G3" score decreases. The shaded area indicates the confidence interval, showing the range of possible scores for each travel time value.
- **G3 by studytime:** This subplot shows an upward trend, indicating that as study time increases, the average "G3" score generally increases. This suggests a positive correlation between study time and performance.
- **G3 by freetime:** This subplot presents a more complex relationship, where the average "G3" score seems to initially increase with free time, reach a peak, and then decrease. This could indicate that a moderate amount of free time might be beneficial for academic performance, but excessive free time may have a negative impact.

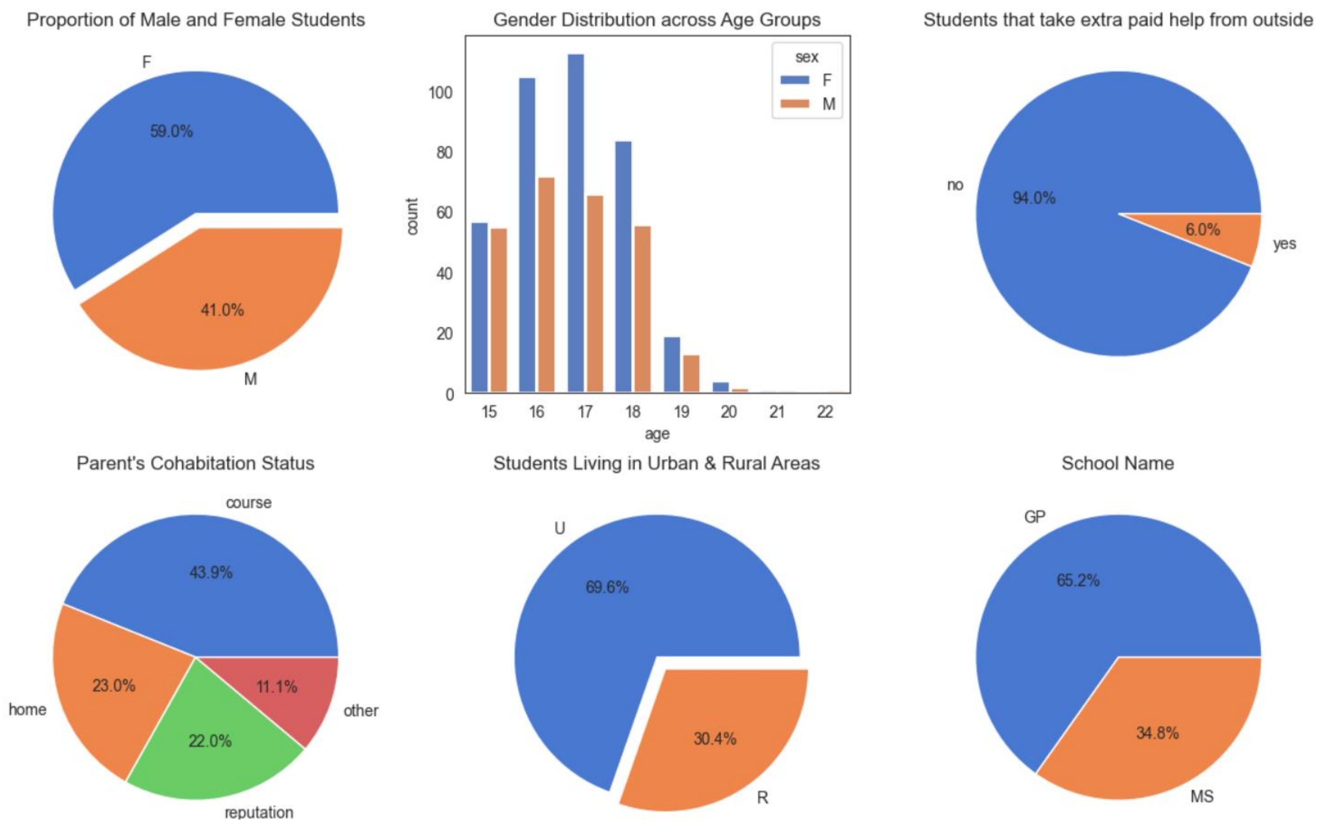


Fig3. Univariate Data Analysis

Insights:

- No. of female students and male students is almost equal, no. of female students is slightly higher.
- Most no. of students are between the ages 15-18
- Most of the students are from Gabriel Pereira School.
- Almost 50% of the students take extra paid classes outside of school.
- The reason majority of the students joined is due to the course.
- Distance from home and school reputation.
- 78% of the students come from Urban areas and 22% from Rural areas.

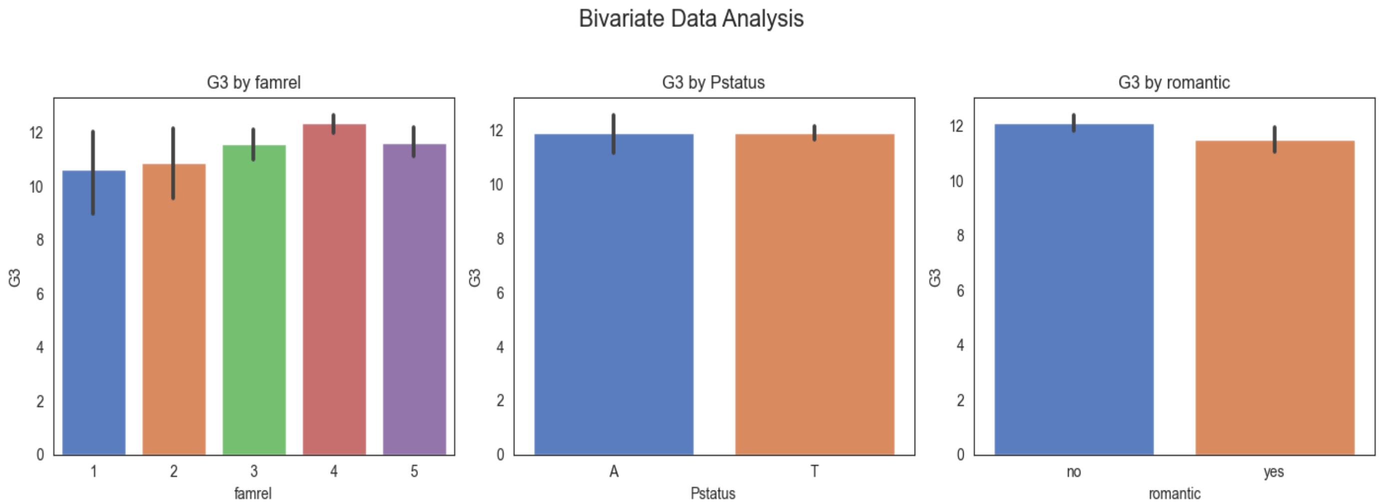


Fig4. Bivariate Data Analysis

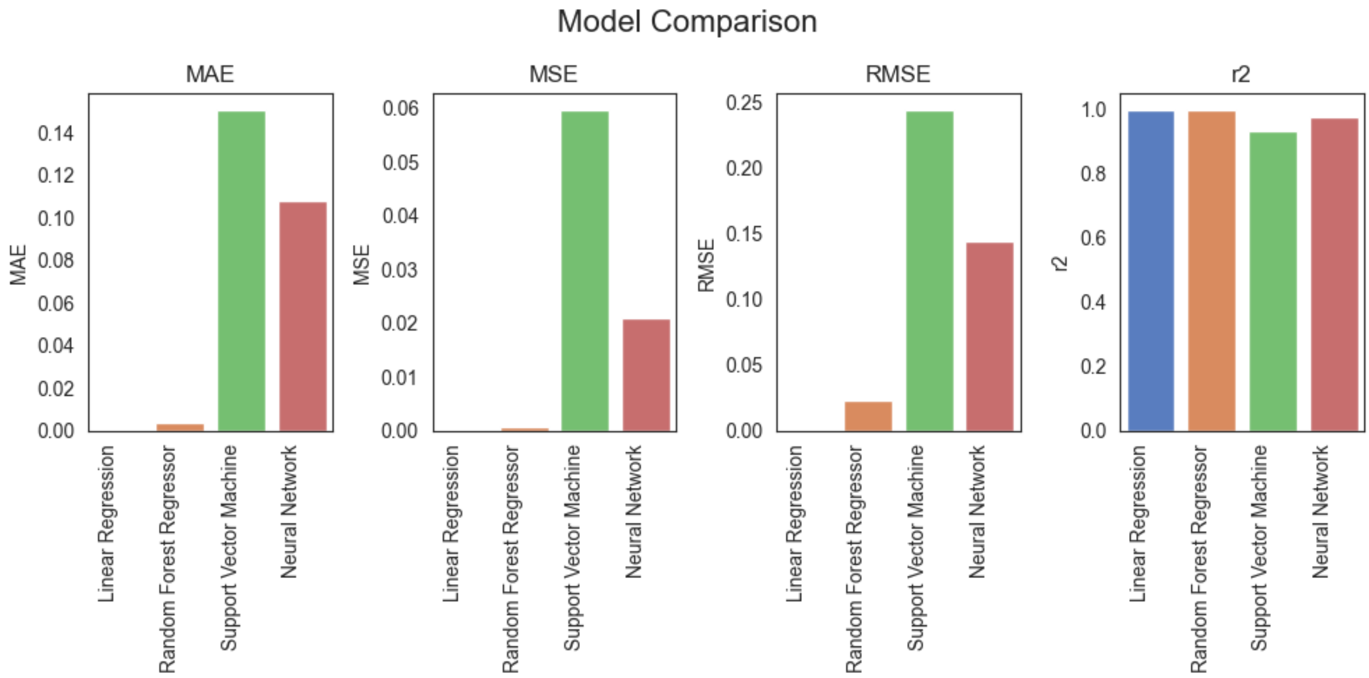
Insights:

- Family of 3 or less show better family relationship.
- Students with Internet availability show higher marks
- Students taking extra paid classes show higher marks.
- Students who study for 5-10 hours a week show higher average marks.
- Optimal free time after school is 2.
    - 1 - very low; 5 - very high
- Consumption of alcohol during the week increases the more they go out with friends.
- Most students scored poorly.

12.2. ML Modelling:

```
# Initialising the models

models = {
    'Linear Regression': LinearRegression(),
    'Random Forest Regressor': RandomForestRegressor(),
    'Support Vector Machine': SVR(),
    'Neural Network': MLPRegressor()
}

# Training the models
for name, model in models.items():
   model.fit(X_train, Y_train)
# Evaluating the models
results = {}
overfit = {}


for name, model in models.items():
   Y_pred = model.predict(X_test)
   mae = mean_absolute_error(Y_test, Y_pred)
   mse = mean_squared_error(Y_test, Y_pred)
   rmse = mean_squared_error(Y_test, Y_pred, squared=False)
   r2 = r2_score(Y_test, Y_pred)


   # Storing results
   results[name] = {'MAE': mae, 'MSE': mse, 'RMSE': rmse, 'r2': r2}
   ### Calculating Overfitting ####
   # Predicting on training data and testing data (Seen and Unseen data)
   training_preds = model.predict(X_train)
   testing_preds = model.predict(X_test)
   # Calculating the MSE
   train_MSE = mean_squared_error(Y_train, training_preds)
   test_MSE = mean_squared_error(Y_test, testing_preds)
   # overfitting values
   overfit[name] = {'Training MSE': train_MSE, 'Testing MSE':test_MSE}
```

Comparison of ML Models:



- Among all the models, Linear Regression model performs well.

➢ **Github link: [Code Implementation]**

13. Conclusion:

This report has presented a compelling case for the development of an AI-powered student performance prediction system, demonstrating its potential to revolutionize educational practices. The proposed system, leveraging a hybrid machine learning model and integrating seamlessly with existing school systems, offers a data-driven approach to personalize learning, identify students at risk, and empower educators, parents, and administrators with valuable insights. The system's ability to predict academic challenges, provide tailored recommendations, and support early intervention strategies holds significant promise for improving student outcomes and fostering a more equitable and effective learning environment. While challenges related to data privacy, bias mitigation, and continuous development remain, the potential impact of this technology on student success and educational excellence makes it a worthwhile investment for the future of learning.

14. References:

El-Sabagh, H.A. Adaptive e-learning environment based on learning styles and its impact on development students' engagement. Int J Educ Technol High Educ 18, 53 (2021).

Ilić, Milena P., et al. "Needs and performance analysis for changes in higher education and implementation of artificial intelligence, machine learning, and extended reality." Education Sciences 11.10 (2021): 568.

Kara, N., & Sevim, N. (2013). Adaptive Learning Systems: Beyond Teaching Machines. Contemporary Educational Technology, 4(2), 108-120.

Owoc, M.L., Sawicka, A., Weichbroth, P. (2021). Artificial Intelligence Technologies in Education: Benefits, Challenges and Strategies of Implementation. In: Owoc, M.L., Pondel, M. (eds) Artificial Intelligence for Knowledge Management. AI4KM 2019. IFIP Advances in Information and Communication Technology, vol 599. Springer, Cham.

National Education Policy 2020:

https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf

Copyright Act, 1957: https://copyright.gov.in/documents/copyrightrules1957.pdf