EduPredict - Big Data Based Student Performance Predictor An AI-Driven Education Analysis System

Anas Junaid 1367063

Muneeb Faisal 1385556

Muhammad Huzaifa Khan1364260

Shaheer khan 1361885

Azmeer Khan 1385447

Usama 1371742

Here is your **complete project report** for **EduPredict – Big Data Based Student Performance Predictor**. You can copy-paste this into Microsoft Word and format it as needed (e.g., apply proper headings, spacing, and alignment).

**EduPredict - Project Report**

**Index**

1. Introduction
2. Objectives
3. Problem Statement
4. Hardware/Software Requirements
5. Functional Requirements
6. Non-Functional Requirements
7. Project Deliverables

**1. Introduction**

The thirst for learning, upgrading technical skills, and applying concepts in real-life environments at a fast pace is what the industry demands from IT professionals today. However, busy work schedules, far-flung locations, and unavailability of convenient time slots pose major barriers when it comes to applying the concepts realistically. Hence, there's a need for alternative implementation means via a laddered approach.

These issues also constrain students. With their packed schedules, it's difficult to keep up with the constant need for integrated application, especially in IT education where technologies evolve rapidly. Thankfully, technology can also provide the solution.

Keeping this in mind, Aptech has revolutionized how students learn and implement concepts by introducing a **live and synchronous eProject learning environment**.

**What is the eProject?**

The eProject is a **step-by-step simulation of classroom and lab environments** for real-life project implementation. It allows students to:

* Practice through a laddered approach.
* Build more robust, larger applications.
* Use real-world utilities in their own designs.
* Convert single programs into unified, complete applications.
* Learn implementation concepts in phases.
* Enhance their technical skills and practical value.
* Work on real-life problems.
* Get mentored through email support.

Students are expected to complete and submit this eProject with documentation and source code to the eProjects Team.

**2. Objectives of the Project**

The objective is to provide a **real-life project scenario** for students to implement using tools they've learned.

* It is not to teach concepts from scratch, but to simulate application.
* Students revise their skills before beginning the project.
* A clear understanding of the subject is essential.
* Students can contact the eProjects Team for clarifications.

**3. Problem Statement**

**Background**

Educational institutions face challenges like:

* Declining student retention
* Inefficient resource allocation
* Lack of personalized learning

**Advancements in Big Data and predictive analytics** offer solutions to these issues.

**Solution: EduPredict**

By leveraging **Hadoop** and **predictive analytics**, EduPredict can:

* Predict student performance
* Detect dropout risks
* Forecast demand for courses
* Provide insights to enhance educational quality

**4. Hardware/Software Requirements**

**Hardware Requirements**

* Minimum Intel i5 (4 Cores), preferably i7
* 16 GB RAM
* 500 GB SSD storage
* Graphics Card
* Windows 10 (64-bit) or higher

**Software Requirements**

* Jupyter Anaconda Notebook 3
* RStudio
* Visual Studio Code / PyCharm
* MongoDB Compass & Shell
* Hadoop, HDFS, Apache Server
* Tableau
* Impala Server

**5. Functional Requirements**

**User Authentication & Role Management**

* Secure login for different user roles: Admin, Teachers, Students, Analysts
* Role-based access controls

**Data Ingestion**

* Import academic records, demographics, LMS data, attendance
* Supports both historical and real-time data
* Compatible with standard data formats

**Data Storage**

* HDFS for scalable, fault-tolerant data storage
* Efficient data partitioning

**Data Processing**

* Parallel data processing for large volumes
* Algorithms to identify trends, anomalies
* Graceful handling of incomplete data

**Real-time Data Handling**

* Integration of real-time streams
* Combined batch + real-time analytics

**Machine Learning Models**

* Predict performance, dropouts, and course demand
* Analyze trends and detect anomalies
* Regularly updated with new data

**Dashboards & Visualization**

* Interactive dashboards using Tableau
* Easy-to-read graphs for insights and patterns

**Alerts & Notifications**

* Threshold-based alerts for performance or anomalies
* Real-time configurable alerts

**Feedback & Support System**

* Users can raise issues or request help
* Continuous support and feedback loop

**6. Non-Functional Requirements**

**Performance**

* Minimal latency for large data volumes
* Real-time monitoring and reporting
* System optimization for efficiency

**Security**

* Encryption for data storage and transmission
* Compliance with data protection laws

**Data Integrity**

* Maintain data accuracy across lifecycle
* Handle missing/incomplete data without failure

**Reliability**

* 99%+ uptime with planned maintenance
* Automated backups to prevent loss

**Scalability**

* Horizontal scaling support
* Load balancing for efficient operations

**Performance Monitoring**

* Continuous tracking of usage and resources
* Integrated with performance tuning tools

**Compliance**

* Follows best practices in Big Data and education analytics
* Uses standard protocols and secure frameworks

**Documentation**

* **User Documentation**: Guides, FAQs, and tutorials
* **Developer Documentation**: System architecture, ML models, workflows
* **Video**: Demonstration video of the entire working system

**7. Project Deliverables**

You are expected to submit a complete project package containing:

* Problem Definition
* Design Specifications
* Flowcharts, Data Flow Diagrams (DFDs)
* Complete Source Code
* Test Data
* Installation Instructions
* Video demonstration
* Optionally: Live hosted link
* A **ReadMe.doc** file listing any assumptions



