**MACHINE LEARNING BASED PREDICTIVE ANALYTICS OF STUDENT PERFORMANCE IN STEM EDUCATION**

1. **SYNOPSIS**

Every government and nation in the world works very hard to develop the education sector since it is a crucial component of society. The educational systems have been impacted by the corona-virus outbreak that has disturbed life in 2020 in many different ways. To get the best results and lower the failure rate, educational systems now need creative strategies to raise the quality of instruction. Machine learning (ML) techniques have gained significant traction recently across a range of global businesses, including education. The term "learning" describes the automatic adjustment or enhancement of an algorithm based on prior "experiences" without any outside aid from a person. The most crucial job in every educational setting is to keep an eye on and raise student performance. This method uses machine learning and data analytics to forecast student success. We can develop a model to help students do better on tests with the aid of the data we have gathered. It is clear that this is a prediction problem, thus we will forecast a student based on the data we have collected and compare different Machine Learning methods, including Naïve bayes, K-nearest neighbor and Support vector machine. This project allows us to ascertain:

1. predict whether a student will pass his final exam or not
2. came up with the best predictive algorithm to calculating the accuracy of the data collected from the college
3. compare which algorithm is best for the students performance prediction
4. **SOFTWARE REQUIREMENTS**

2.1 Anaconda:

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single machine. It can be easily installed on any OS such as Windows, Linux, and MAC OS. It provides more than 1500 Python/R data science packages which are suitable for developing machine models.

2.2 Jupyter Notebook:

In the past couple of years, Notebooks have become a popular tool in fields like data science and machine learning. It is a web-based interactive computing platform. The notebook combines live code, equations, narrative text and visualizations.They're used a lot in machine learning, mainly for experimentation and visualization.

1. **MODULE DESCRIPTION**

3.1 Python:

Python is a [high-level](https://en.wikipedia.org/wiki/High-level_programming_language), [general-purpose programming language](https://en.wikipedia.org/wiki/General-purpose_programming_language). Its design philosophy emphasizes [code readability](https://en.wikipedia.org/wiki/Code_readability) with the use of [significant indentation](https://en.wikipedia.org/wiki/Off-side_rule). Python is [dynamically typed](https://en.wikipedia.org/wiki/Type_system#DYNAMIC) and [garbage-collected](https://en.wikipedia.org/wiki/Garbage_collection_(computer_science)). It supports multiple [programming paradigms](https://en.wikipedia.org/wiki/Programming_paradigm), including [structured](https://en.wikipedia.org/wiki/Structured_programming) , [object-oriented](https://en.wikipedia.org/wiki/Object-oriented_programming) and [functional programming](https://en.wikipedia.org/wiki/Functional_programming). It is often described as a "batteries included" language due to its comprehensive [standard library](https://en.wikipedia.org/wiki/Standard_library).

* 1. Naïve Bayes:

Naïve bayes is a probabilistic classifier used in machine learning that applies the Bayes theorem with strong (nave) independence assumptions between the features. Because the future grades of the specific course chain must be predicted if the specifics of earlier courses or subjects are known, the Nave Bayes Classifier is chosen. One of the effective and efficient inductive learning methods for machine learning is naive bayes.

3.3. K-nearest neighbor:

It is a type of instance-based learning, or lazy learning, in which the complete computation is postponed in expectation of classification and the function is just roughly calculated nearby. The primary issue with the k-nearest neighbor method is that loud or unsuitable characteristics can severely impair its accuracy.

* 1. Support Vector Machine:

Support The idea of planes is part of the vector machine. A plane will have groups of unique instances from various classes. It uses various classification methods. The simplest is linear, which divides sets of instances into their groups using nothing more complicated than a line. More complicated structures are utilized for classification that is more precise. We have employed a variety of kernel functions in our classification model, including linear, polynomial, and gaussian kernels. The addition of classes during the classification phase may lower the SVM's success rate. Nonetheless, it can be used to solve issues in an efficient manner.