PRIOR KNOWLEDGE

Team ID	PNT2022TMID01667
Project Name	Project - Smart Lender - Applicant Credibility Prediction for Loan Approval

MACHINE LEARNING

Machine learning is a subfield of artificial intelligence (AI) and computer science that utilises data and algorithms to mimic how people learn, progressively improving its accuracy.

Machine learning is a critical component of the rapidly expanding discipline of data science. Algorithms are taught using statistical approaches to produce classifications or predictions and to find critical insights in data mining operations. These insights then influence decision making within applications and enterprises, ideally influencing key growth indicators. As big data expands and grows, so will the market demand for data scientists. They will be expected to assist in identifying the most pertinent business questions and the data to address them.

TensorFlow and PyTorch are two frameworks used to expedite solution creation while developing machine learning algorithms.

ALGORITHMS USED

DECISION TREE

Decision trees may be used to forecast numerical values (regression) as well as categorise data. A decision tree is a branching series of related choices represented by a tree diagram. One of the benefits of decision trees over neural networks is that they are easier to evaluate and audit.

TYPES OF DECISION TREE:

ID3 : Ross Quinlan is credited with creating ID3, which stands for "Iterative Dichotomiser 3." To evaluate potential splits, this technique uses entropy and information gain as metrics

C4.5 : This algorithm is a subsequent iteration of ID3, which was also created by Quinlan. It may assess split points inside decision trees using information gain or gain ratios.

CART: CART stands for "classification and regression trees," and it was coined by Leo Breiman. This method often makes use of Gini impurity to choose the best characteristic to split on. Gini impurity quantifies how frequently a randomly selected characteristic is misclassified. When assessing Gini impurity, a lower value is preferable.

RANDOM FORESTS

In a random forest, the machine learning algorithm predicts a value or category by combining the results from a number of decision trees.The random forest algorithm is a bagging technique extension that use both bagging and feature randomization to produce an uncorrelated forest of decision trees. Feature randomization, also known as feature bagging or "the random subspace approach" which provides a random selection of features, ensuring minimal correlation among decision trees. This is a significant distinction between decision trees and random forests. Random forests, in contrast to decision trees, only choose a subset of the potential feature splits.

K-NEAREST NEIGHBORS

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another

XGBOOST

XGBoost, or Extreme Gradient Boost, is a machine learning technique used to create gradient boosting decision trees. When it comes to unstructured data, such as photos and unstructured text data, ANN models (Artificial neural network) appear to be at the top of the list when it comes to prediction. When it comes to structured/semi-structured data, decision trees are currently the best.

TECHNOLOGIES USED

PYTHON

Python is an agile, dynamically typed, expressive, open source programming language that supports multiple programming philosophies,including procedural, object-oriented, and functional.Python is a popular high-level programming language that is easily extensible through the use of third-party packages and often allows powerful function to be written with few lines of code.

HTML

HTML is a programming language used to organise web content. Its goal is to make online design and development easier by developing a standardised and user-friendly UI markup language. HTML allows you to deconstruct and compartmentalise your pages, as well as build separate components that are not only meant to arrange your site logically, but are also designed to provide your site syndication capabilities. HTML might be referred to as the "information mapping method to website design" since it integrates the concept of information mapping, separating and labelling data to make it easier to use and comprehend. This is the basis for HTML tremendous semantic and graphical usefulness.

CSS

CSS is an abbreviation for Cascading Style Sheets. It is a style sheet language that is used to specify the look and formatting of a markup

language document. It adds a new functionality to HTML. It is commonly used in conjunction with HTML to alter the appearance of web pages and user interfaces. It may also be used with any type of XML document, such as plain XML, SVG, or XUL.

JAVASCRIPT

JavaScript (js) is a lightweight object-oriented programming language that is used to script webpages on many websites. When applied to an HTML document, it is an interpreted, full-fledged programming language that enables dynamic interactivity on website. Users may use JavaScript to create contemporary web applications that allow them to engage without having to reload the page every time. The typical website makes use of js to give various levels of interactivity and simplicity**.**

TOOLS USED

FLASK

Flask is a terrific and relatively light Python framework for constructing sophisticated online apps, as well as an amazing tool for generating dynamic and interactive webpages. You can construct fascinating applications using this strategy, even on versions before to 7.X. We can acquire outstanding results quickly using Flask and update management procedures with new apps.

IBM CLOUD

The term "IBM Cloud" refers to IBM's hardware, software, and services for assisting organisations in building private clouds, as well as its Bluemix public cloud services. The moniker "Bluemix" was formerly reserved for IBM's PaaS services for developers, however Bluemix now now provides certain IaaS services.