**Smart Investment Prediction**

**Introduction:**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed**. Machine learning focuses on the development of computer programs** that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. **The primary aim is to allow the computers learn automatically** without human intervention or assistance and adjust actions accordingly.

Some machine learning methods:

Machine learning algorithms are often categorized as supervised or unsupervised.

1. ***Supervised machine learning*** algorithmscan apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.
2. ***Unsupervised machine learning*** algorithmsare used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.
3. ***Reinforcement machine learning*** algorithmsis a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

**Objective of the research:**

An investment is made because it serves some objective for an investor. Depending upon the life stage and risk appetite of the investor, there are three main objectives of investment: safety, growth and income. Every investor invests with a specific objective in mind, and each investment has its own unique set of benefits and risks. Let us understand these objectives in detail.

1. Safety:

While no investment option is completely safe, there are products that are preferred by investors who are risk averse. Some individuals invest with an objective of keeping the money safe, irrespective of the rate of return they receive on their capital. Such near-safe products include fixed deposits, savings accounts, government bonds etc.

1. Growth:

While safety is an important objective for many investors, a majority of them invest to receive capital gains, which means that they want the invested amount to grow. There are several options in the market that offer this benefit. These include stocks, mutual funds, gold, property, commodities, etc. It is important to note that capital gains attract taxes, the percentage of which varies according to the number of years of investment.

1. Income:

Some individuals invest with the objective of generating a second source of income. Consequently, they invest in products that offer returns regularly like bank fixed deposits, corporate and government bonds, etc.

Some people invest their money in various financial products solely for reducing their tax liability. Some products offer tax exemptions while many offer tax benefits on long-term profits.

1. Liquidity:

Many investment options are not liquid. This means they cannot be sold and converted into cash instantly. However, some people prefer investing in options that can be used during emergencies. Such liquid instruments include stock, money market instruments and exchange-traded funds, to name a few.

Other objectives:

The objectives mentioned above are the most common ones for today. Some other objectives include:

1. Tax exemption:

The main objective of this research is to know whether to invest money in a company or not. It is based upon the prediction values 0 and 1. If the prediction value is zero, then it tells you not to invest the money. If the prediction value is one, then it tells you to invest money.

Problem statement:

Problem statement is to find whether to invest money or not in a company based upon the company’s grading, share prices and the number of projects the company taken.

So, by analyzing these values, we will be able to know whether to invest or not.

Industry profile:

Companies in this industry take equity positions in target companies, typically owning a controlling interest. Major companies include Berkshire Hathaway, Clark Enterprises, Icahn Enterprises, and Jeffries Financial Group (all based in the US), along with CITIC (Hong Kong), EXOR (Italy), and SoftBank (Japan). The industry also includes the investment arms of major banks, such as JP Morgan Chase, Goldman Sachs, Citigroup, Bank of America, and Deutsche Bank.

Activity in emerging markets in the Asia/Pacific region has attracted the interest of companies seeking new investment opportunities. Political uncertainties in some emerging regions, such as the Middle East, can pose risks that outweigh potential returns and cause investors to look elsewhere.

The US investment firm industry includes about 6,900 establishments (single-location companies and units of multi-location companies).

**Competitive landscape:**

Demand is driven by the movement of capital throughout the economy. Profitability depends on the ability to identify targets for investment. Large firms have advantages in **access to capital.** Small firms can compete by specializing in market segments.

Investment firms compete with other entities seeking investment capital. Competing organizations may include sponsors of public and private investment funds, **real estate development companies**, and other organizations within the global investment management industry.

**Products, operations & technology:**

The main revenue sources for investment firms are **interest and dividends** (about 70%). Other sources include sales of goods and services, administrative and management services fees, and capital gains from asset sales.

Data collection:

The dataset for this model can be checked through the below link:

[Smart Investment Prediction.csv](file:///E:\Smart%20Investment%20Prediction.csv)

Features in the dataset:

* Grade of the company
* No. of offline projects
* No. of client projects
* Net turnover
* Share price

The important features are grade, no. of offline projects, no. of client projects, share price through which we can predict the output.

Figures and tables:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Serial no. | Grade of the company | No. of client  projects | No. of offline  projects | Net turnover | Share price |
| 1. | 1 | 85 | 256 | 1265242 | 250.16 |
| 2. | 1 | 85 | 256 | 1265243 | 250.15 |
| 3. | 1 | 85 | 256 | 1265244 | 250.14 |
| 4. | 1 | 85 | 256 | 1265245 | 250.13 |
| 5. | 2 | 92 | 261 | 1625345 | 425.3 |

Table: Investment details of five companies

Statistical techniques and visualization:

There are different types of classification models. They are:

1. Logistic Regression:

It is a statistical method for analyzing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables.

#### **Decision Trees:**

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches and a leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

#### **Random Forest:**

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees’ habit of over fitting to their training set.

#### **Nearest Neighbor:**

The k-nearest-neighbors algorithm is a classification algorithm, and it is supervised: it takes a bunch of labelled points and uses them to learn how to label other points. To label a new point, it looks at the labelled points closest to that new point (those are its nearest neighbors), and has those neighbors vote, so whichever label the most of the neighbors have is the label for the new point (the “k” is the number of neighbors it checks).

1. SVM classifier:

A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are more commonly used in classification problems and are based on the idea of finding a hyperplane that best divides a dataset into two classes.

Among all these classifier models, this model is developed through decision tree classifier because this model of classification has the highest accuracy when compared to other models.

Data modelling and visualization:

A decision tree is a flowchart-like tree structure where an internal node represents feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret.

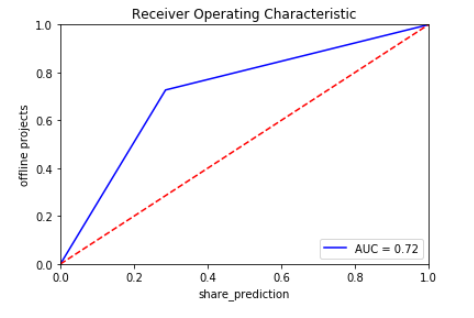


Figure: ROC curve for decision tree classifier

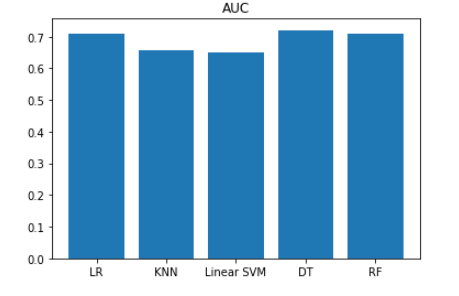


Figure: AUC curve of all the models

Conclusion:

This model enhances the people to make the exact decision on the problem of investing their money in a company because the investment among the people of a country is an essential prerequisite for capital formation and the faster growth of an economy. Investment culture refers to the attitudes, perceptions, and willingness of the individuals and institutions in placing their savings in various financial assets, more popularly known as securities.