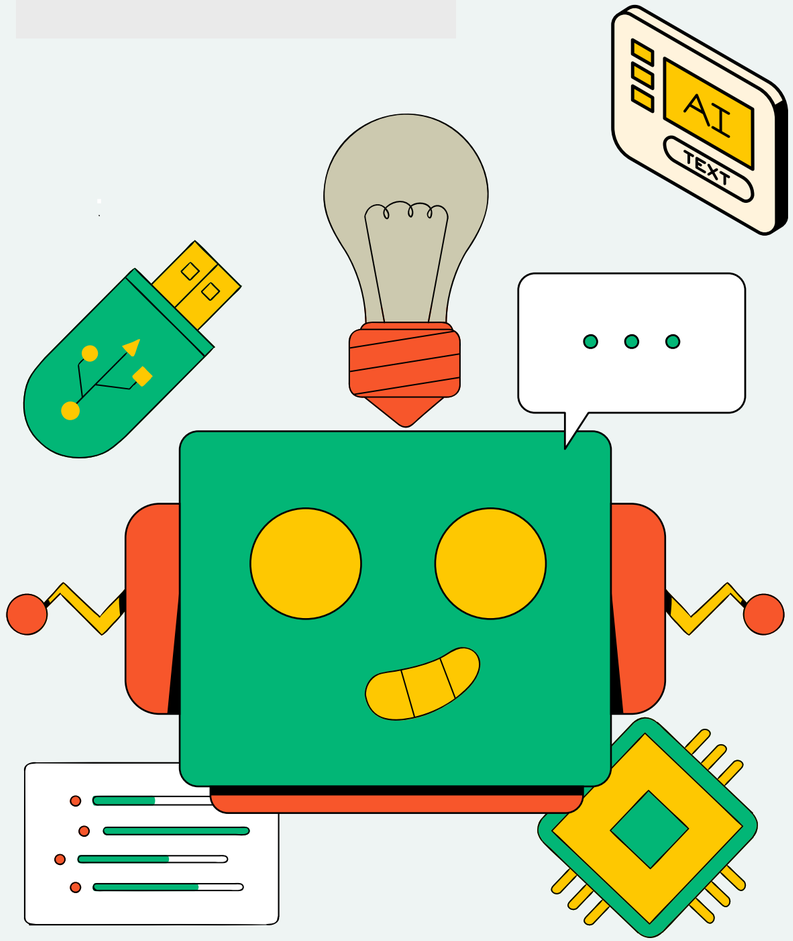


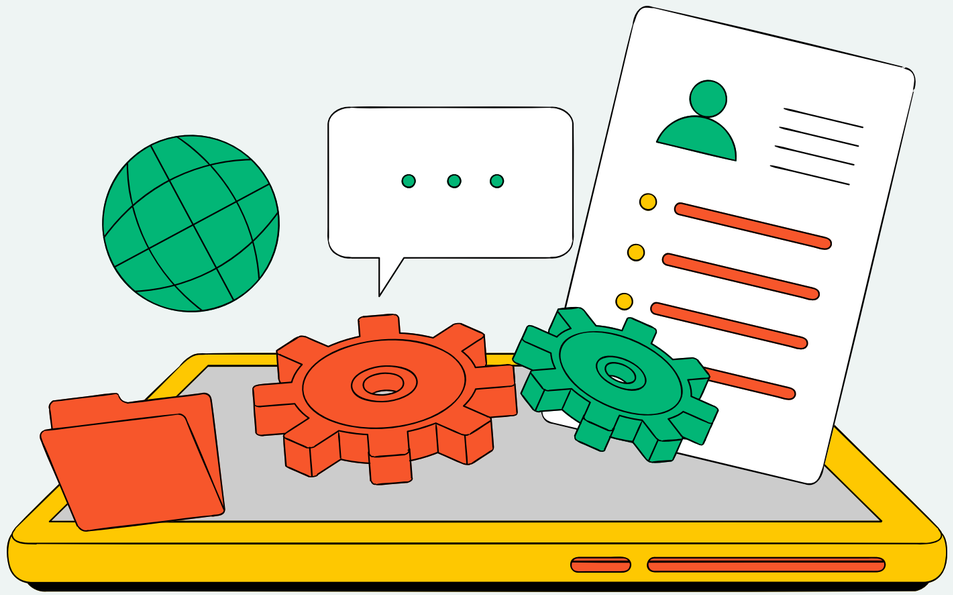
SUPERVISED LEARNING

TYPES OF MACHINE LEARNING

Supervised Learning

Unsupervised Learning

Reinforcement Learning



SUPERVISED MACHINE LEARNING

Supervised learning uses labelled data to train a model for predictions by providing input features and output values for decision-making.

Supervised learning is a machine learning technique that is widely used in various fields such as finance, healthcare, marketing, and more. It is a form of machine learning in which the algorithm is trained on labeled data to make predictions or decisions based on the data inputs.

WHAT IS SUPERVISED LEARNING?

For example, assume that a machine has to predict whether a customer will buy a specific product let's say "Antivirus" this year or not.

The machine will do it by looking at the previous knowledge/past experiences i.e. the data of products that the customer had bought every year and if he buys an Antivirus every year, then there is a high probability that the customer is going to buy an antivirus this year as well. This is how machine learning works at the basic conceptual level.

STANDARD SUPERVISED LEARNING ALGOS

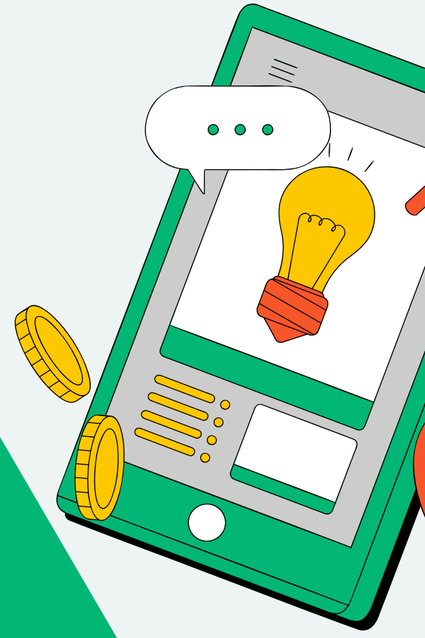
Decision Trees

K nearest neighbours

Random forests

Support nearest vectors

Logistic regression

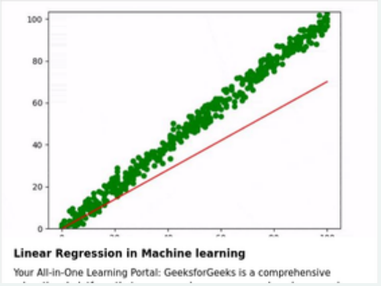


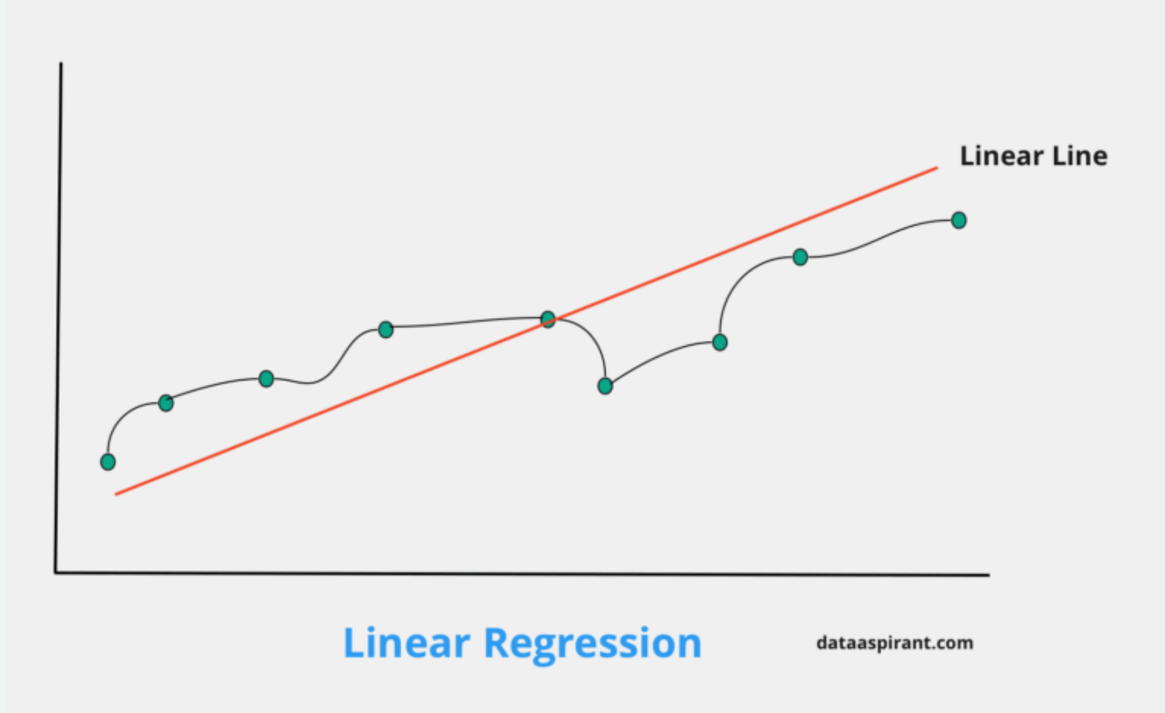
TASKS WHERE SUPERVISED LEARNING IS APPLIED

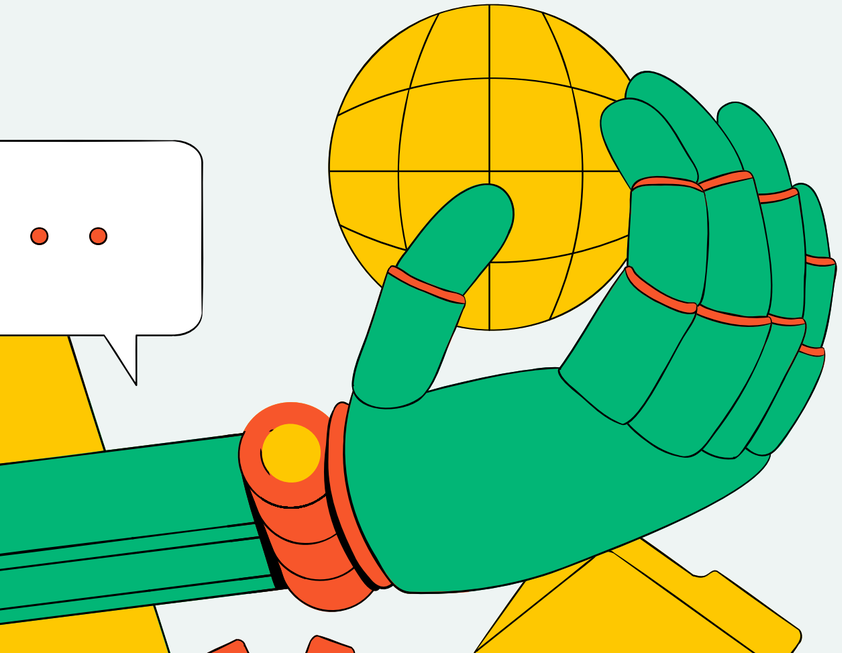
- Image and object recognition
- Spam filtering
- Sentiment analysis
- Weather forecasting

LINEAR REGRESSION

It is commonly used to predict the value of a continuous variable from the input given by multiple features or to classify observations into categories by considering their correlation with reference variables.

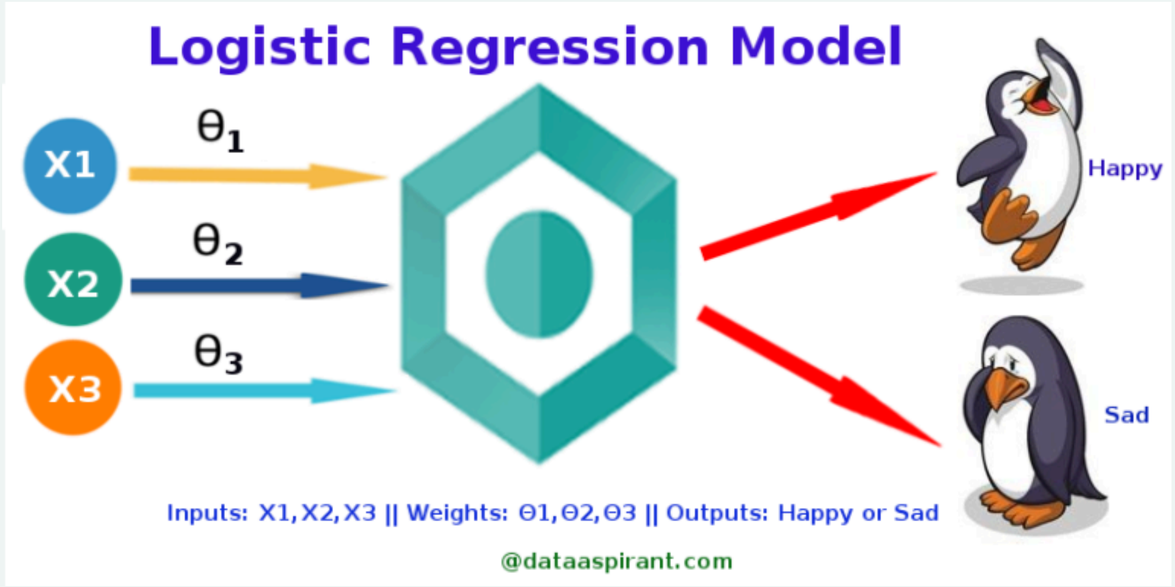


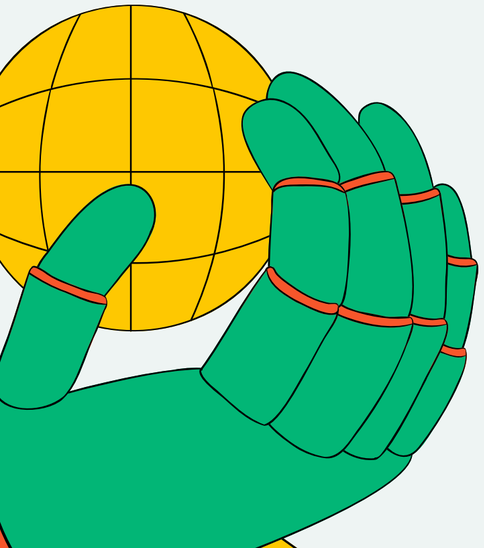




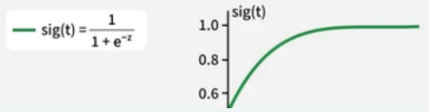
LOGISTIC REGRESSION

Logistic regression is a type of predictive modelling algorithm used for classification tasks. It is used to estimate the probability of an event occurring based on the values of one or more predictor variables.





Content is no longer available



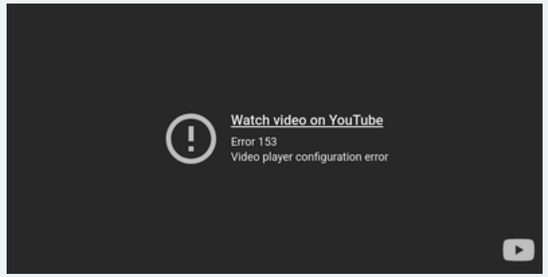
<u>Logistic regression models</u> are commonly used for binary classification problems, such as predicting whether a customer will subscribe to a service. However, logistic regression can also be used for <u>multiclass classification</u> or prediction tasks.

Advantage: simplicity and interpretability.

SUPPORT VECTOR MACHINES (SVM)

Support Vector Machines (SVM) are robust algorithm that use a kernel to map data into a high-dimensional space and then draw a linear boundary between the distinct classes.

SVMs are often used for text classification or image recognition because they can accurately predict categorical variables from large datasets.



1.4. Support Vector Machines

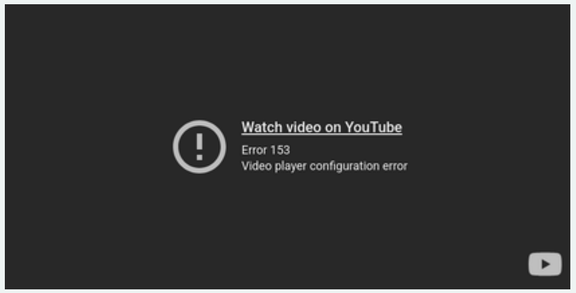
Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. The advantages of support vector machines are: Effective in high ...

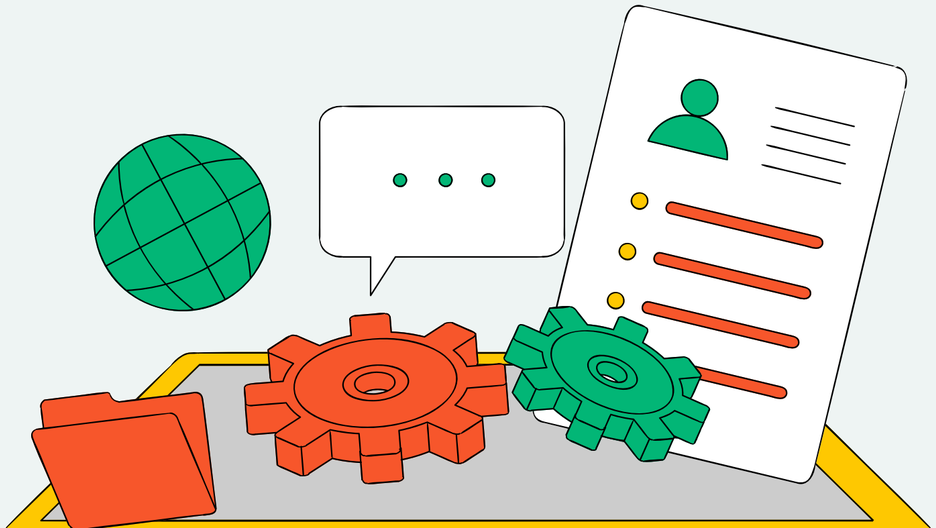
scikit-learn

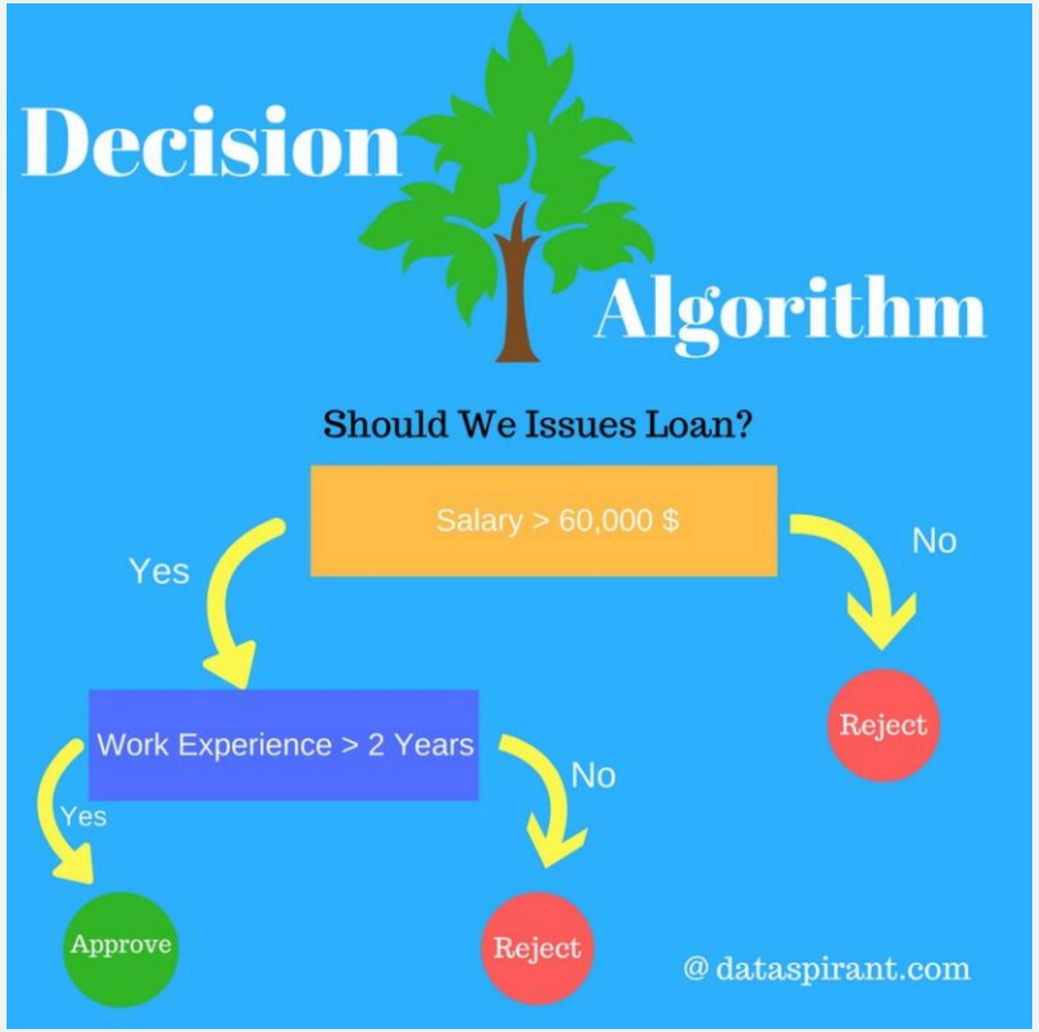
Nevertheless, SVMs have several advantages over other supervised learning algorithms. For instance, they are highly tolerant of noise and can handle non-sense features. Furthermore, SVMs effectively handle data sets with large feature vectors by exploiting the kernel trick. Additionally, they work well when there is a significant overlap between classes because SVM is less sensitive to outliers than other models, such as decision trees or linear regression.

DECISION TREES

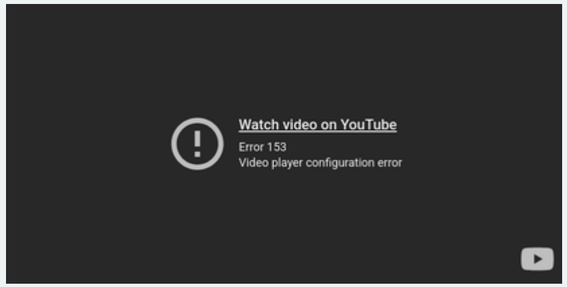
Decision Trees are popular supervised learning algorithms for classification and regression tasks using a "tree" structure to represent decisions and their associated outcomes.



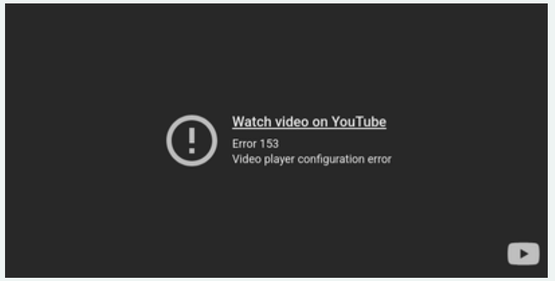




RANDOM FORESTS



Random forests are essentially multiple decision trees combined to form one powerful "forest" model with better predictive accuracy than individual trees.

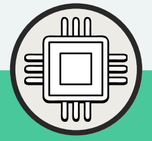


They work by randomly choosing sets of observations as well as variables at each node split within the forest, comparing several predictive models before coming up with an optimal outcome that maximizes accuracy.
With the ability to easily see which variable is responsible for each prediction using decision trees, they are easy to analyze and interpret.

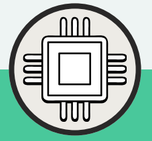
Using random forests model also reduces overfitting due to the fact that many trees are generated, meaning that it's less likely to fit the data too well. In other words it's handles bias and variance well.

As a result, random forests perform better with larger and more complex datasets making it an ideal algorithm for many supervised learning tasks.

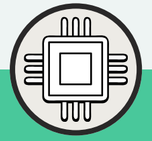
HOW TO IMPROVE PERFORMANCE?



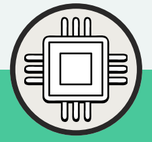
Holdout test data set



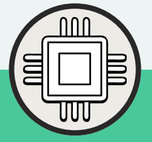
Hyperparameter optimization.



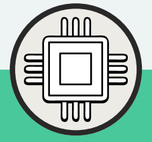
Data preprocessing techniques



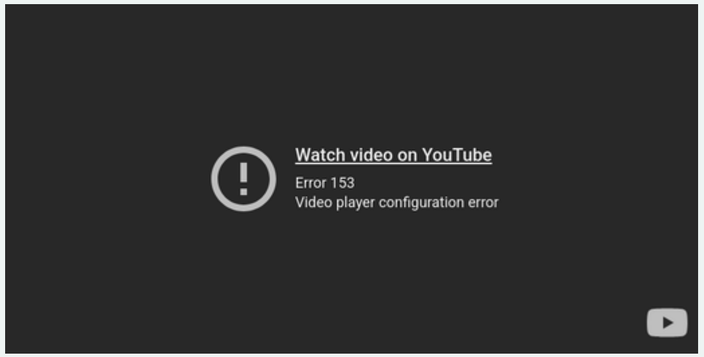
Dataset observation



Feature Engineering



Synthetic Data



k-nearest-neighbor-algorithm-for-machine-learning
Click here to visit the page.

Channels – statquest