# Data science

PRODUCT DEMAND PREDICTION

Project Procedures

- Data collection
- Data pre-processing
- Feature selection
- Model selection
- Training the model
- Validating and testing
- Deployment

- Feedback loop
- Demand planning
- Scenario analysis
- Collaboration
- Hyper parameter tuning
- Monitoring and updating

Data collection

Data collection is the process of gathering and measuring information on variables of interest

Usage data

Surveys

Focus groups

Data collection

Observation

Online tracking

Interviews
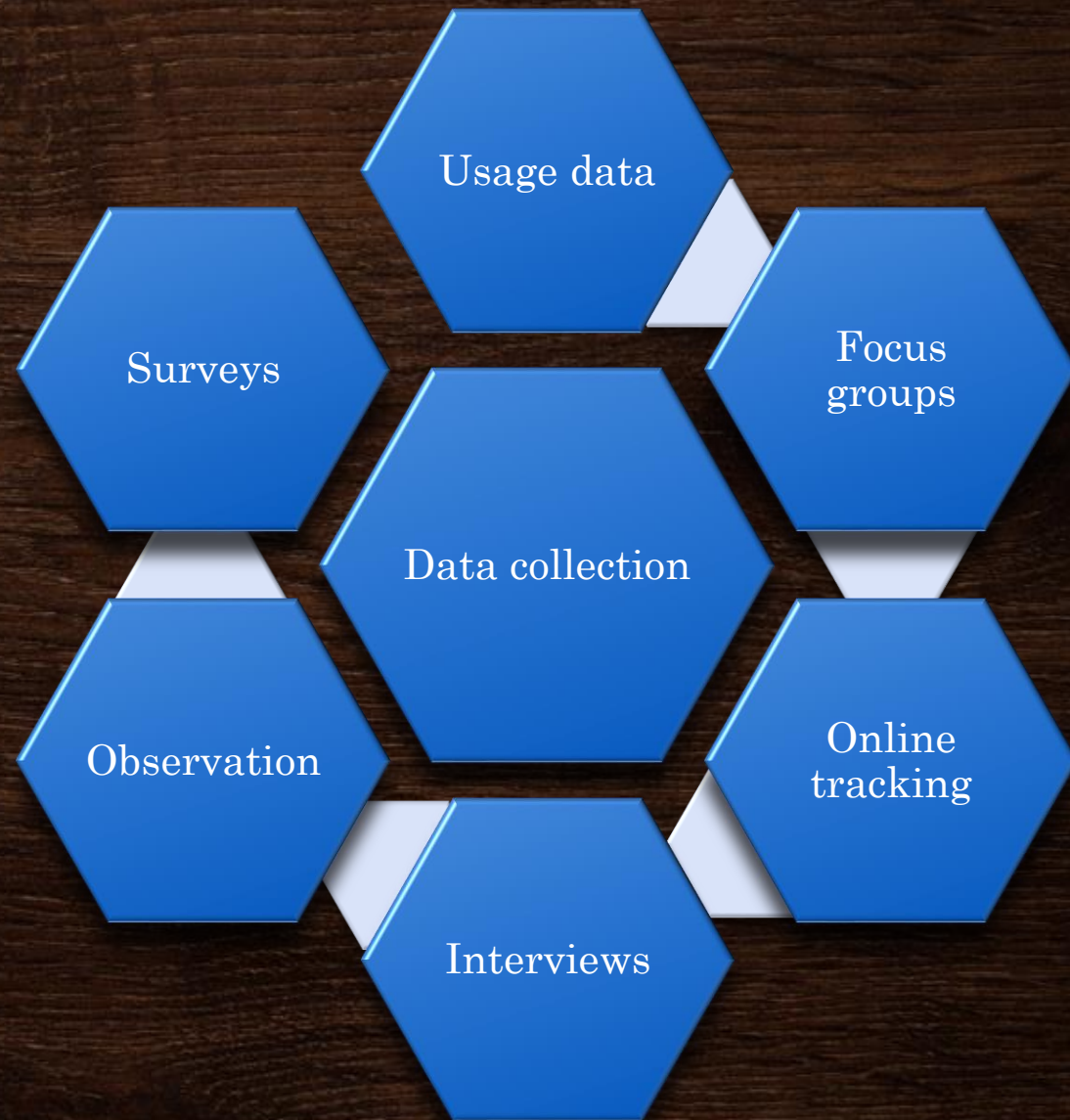
Data pre-processing

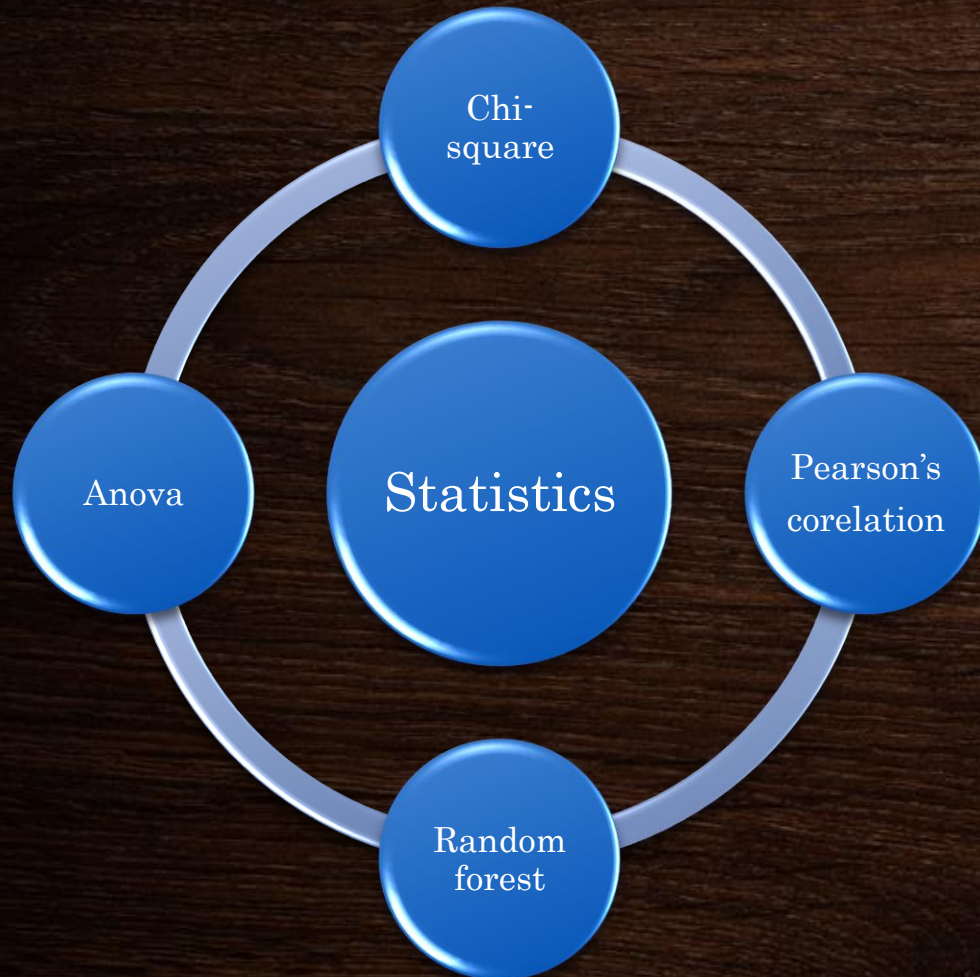The concept of changing the raw data into a clean data set

STEPS

- Data quality assessment
- Data cleaning
- Data transformation
- Data reduction

Feature selection

The method of reducing the input variable to your model by using only relevant data and getting rid of noise in data.

Chi-square

Anova

Statistics

Pearson's corelation

Random forest

ADVANTAGES

- Data understanding
- Model improvement
- Model interpretability

Model selection

The process of choosing one among many candidate models for a predictive modelling problem

Mode selection based by

- Performance
- Explainability
- Complexity
- Dataset size
- Dimensionality
- Training time and cost
- Inference time
- Conclusion

EXAMPLES

- Logistic regression
- Random forest
- SVM
- XGBoost
- Naïve Bayes

XGBoost

XGBoost Algorithm

XGBoost is an optimized distributed gradient boosting library designed for efficient and scalable training of machine learning models.

STEPS

- Install XGBoost
- Import the necessary files
- Load and prepare your data
- Split data into Training set and testing set
- Initialize and train XGBoost model
- Make predictions
- Evaluate the model
- Hyper parameter tuning
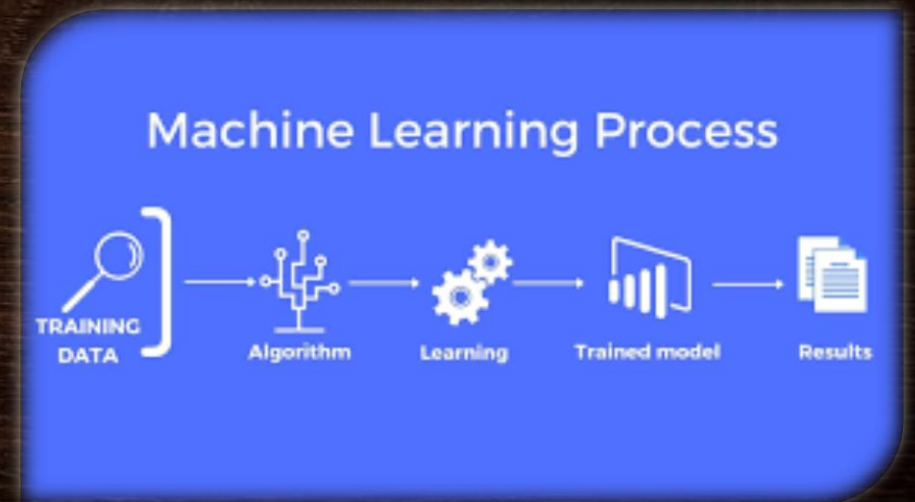- Feature importance
- Final Visualization
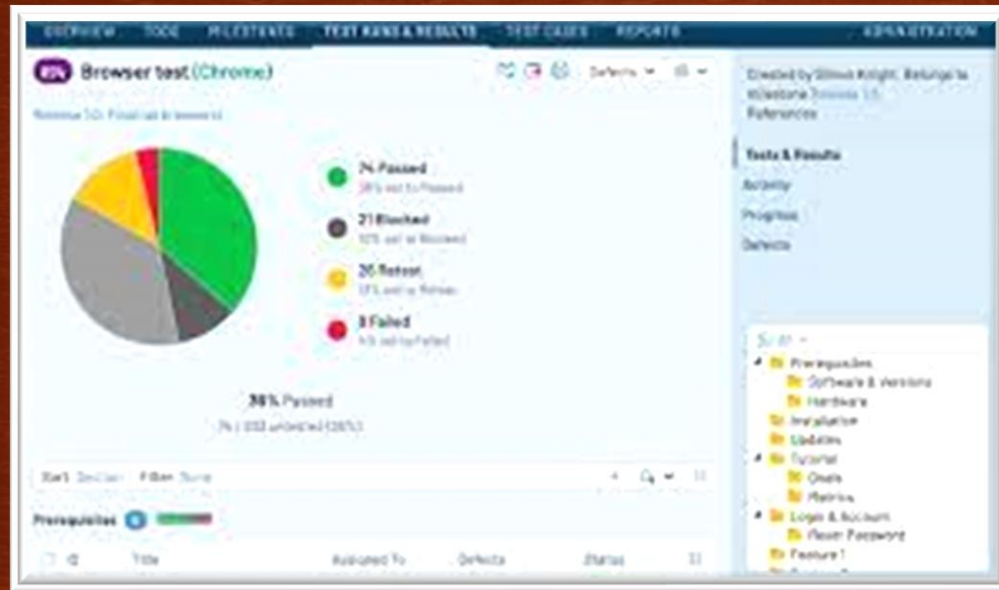
xgboost

Training the model

learning (determining) good values for all the weights and the bias from labeled examples

How to train

- Begin with existing data
- Analyze data to identify patterns
- Make predictions

Validating and testing

The static practice of studying and verifying the specific requirements of a particular stage in development

DEPLOYMENT

the process of putting machine learning models into production

USES OF DEPLOYMENT

The process by which the document types, model, and overall project definition are made available for use.
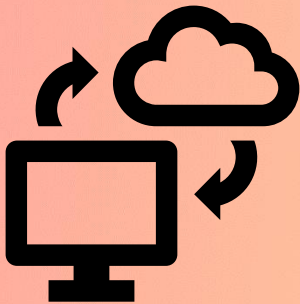
Feedback loop

a process in which the outputs of a system are circled back and used as inputs

ELEMENTS OF FEEDBACK LOOP

- Stimulus
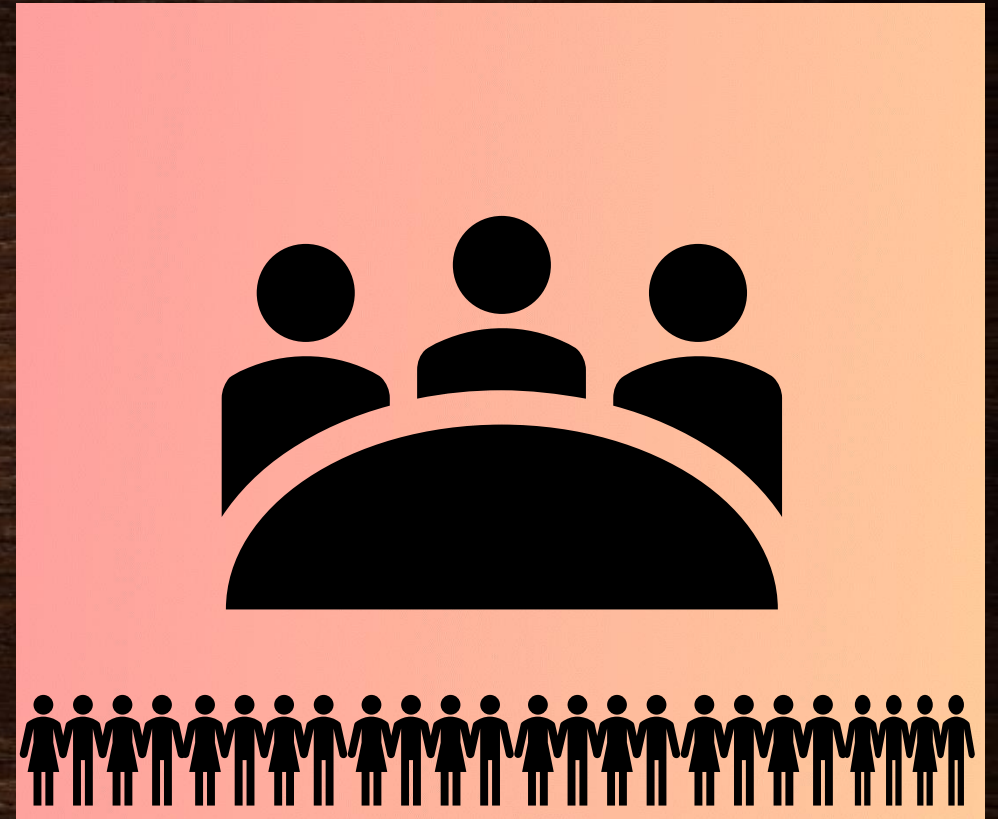- Sensor
- Control
- Effector

Demand planning

A supply chain management process of forecasting, or predicting, the demand for products to ensure they can be delivered and satisfy customers

Collaboration

Collaboration entails so much more than data scientists working together on a project

Monitoring and updating

The process of utilizing permanently mounted sensors to regularly check the condition of a machine

The process of tracking the behaviour of a deployed model to analyze performance

Conclusion