Data Exploring

Analyze and investigate data sets and summarize their main characteristics

Answers the question: What Happened or What is Happening?

Exploratory Data Analysis(EDA)

Improve understanding of variables by extracting averages, mean, minimum, and maximum values, etc.

What we use Clustering Separates similar data points into intuitive groups Answers the question: What will they be interested in? K-means Simple to implement, Scales to large datasets, Cluster Unlabeled data Gaussian Process Model Cluster unlabeled data

Find unusual

occurrences

Predict between

categories

Regression

Makes forecasts by estimating the relationship between values

Answers guestions like: How much or how many?

Logistic Regression	Fast training times, linear model
Decision Tree	Accurate, fast training, large memory footprint
Random Forest	Automate Missing Values, Works with both Categorical and Continuous value

Prescriptive Analysis

The use of technology to help businesses make better decisions through the analysis of raw data.

Answers the question: What Should be done? or What can we do to make something happen?

Decision Rules	A simple IF-THEN statement consisting of a condition (also called antecedent) and a prediction.
What-if scenarios	Look at multiple potential actions and relays the consequences of each.
Stochastic Simulation	Easy to talk quantitatively about ranges and likely outcomes.
Markov Decision Process	Modeling decision making in situations where outcomes are partly random and partly under the control of a decision maker

Dimensionality Reduction

Identify strength of

relationship

Predict value

Generate Suggestions

Component

Analysis (PCA)

Identify the strength of the relationships and reduce the number of input variables in a dataset.

Answers the question: What Happened ? or What is Happening?

Scatter Plot	Shows the relationship between two variables , range of data flow , easy to plot the diagram
Principle	mathematical technique for reducing the

mathematical technique for reducing the dimensionality of data, while keeping as much variation as possible.

Classification

LSTM

One class SVM

Anomalies Detection

Identifies and predicts rare or unusual data points

Answers the question: Is this weird?

long-term dependencies

dimensional spaces

Good for time series, Capable of learning

Memory efficient, Effective in high

Answers questions with multiple possible answers

Answers questions like: Is this A or B or C or D?

The state of the s	
One class SVM	Memory efficient , Effective in high dimensional spaces
Logistic Regression	Fast training times, linear model
Naïve Bayes	Easy to implement , Handles both continuous and discrete data , Fast
K-Nearest Neighbors(KNN)	Accurate , Simple to interpret , Quick Calculation time
Decision Tree	Accurate, fast training, large memory footprint
Random Forest	Automate Missing Values , Works with both Categorical and Continuous values