



INSAID

INDUSTRY SPECIFIC PROJECTS





DATA ANALYSIS BASICS WITH PYTHON & EDA



Indian Premier League 2008-2018

Analyse key success factors for top cricket team at IPL

The matches dataset contains 18 variables and 600+ observations . The deliveries dataset contains 21 variables and 160k+ observations of the IPL 2018 season.



Car Sales Advertisement

Analyse the car sale data in Ukraine.

The dataset contains 10 variables and 9k+ observations of the car sales data in Ukraine.



Olympic 1896- 2014

Analyse which country have won the most medals at Olympic games.

The dataset contains 9 variables and 31.2k observations of the summer olympic games (1896 - 2014)



1000 movies data

Analyse the IMDB 1000 most popular movies and come up with interesting insights.

The dataset contains 12 variables and 1000 observations of the top 1000 popular movies for past 10 years



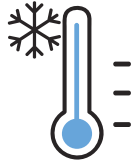
Restaurant across America data

Analyse the availability and accessibility of food across America

The dataset contains 10 variables and 10k observation of the fast food restaurants in America.



MACHINE LEARNING FOUNDATION



Predicting temperatures in World War II

Predict the maximum temperature when minimum temperature is known.

The dataset contains 31 variables and 100k+ observations of the weather conditions during World War



Candy Classification

Classify if a candy is a chocolate or not based on its features

The dataset contains 13 variables and 86 observations to predict the most or least popular Halloween candy.



Predicting Housing prices

Predict final sales price of each house of residential homes in Iowa.

The train dataset contains 81 variables and 1461 observations. The test dataset contains 80 variables and 1460 observations of the house. prices at any anonymous location.



Predicting risk in Life insurance

Develop a predictive model that accurately classifies risk, impacting public perception of the industry.

The train dataset contains 128 variables and 59.4k observations. The test dataset contains 127 variables and 19.8k observations of the insurance applicants.



Credit Card Fraud Detection

Identify the fraudulent credit card transactions.

The dataset contains 31 variables and 285k observations of transactions made by credit cards in September 2013 by European cardholders.



MACHINE LEARNING INTERMEDIATE



SMS Spam collection data

Classify collection of spam messages tagged as spam or legitimate.

The dataset contains 5 variables and 5572 observations collected for SMS spam research.



Simplified Human Activity

Analyse the recordings of subjects performing activities while carrying inertial sensors.

The test dataset contains 562 variables and 1542 observations. the train dataset contains 563 variables and 3609 observations of subjects performing activities while carrying inertial sensors.



Gender recognition by voice

Identify a voice as male or female(SVM)

The dataset contains 21 variables and 3k+ observations to identify a voice as male or female using acoustic properties of voice and speech.



Store Item Demand Forecasting

Predict 3 months of item sales at different store.

The test dataset contains 4 variables and 45000 observations. the train dataset contains 4 variables and 90k observations of a 5 year of store-item-sales data.



Safe driver prediction

Predict the probability that an auto insurance policy holder files a claim.

The test dataset contains 58 variables and 80k+ observations. the train dataset contains 59 variables and 60k+ observations of driving data.



MACHINE LEARNING ADVANCED



Apriori Algorithm(Market Basket Analysis)

Market Basket Analysis of e-commerce data of transaction of 2010 and 2011

The dataset contains 8 variables and 542k observations of all the transaction of 2010 and 2011 for a UK based and registered non - store online retail.



MovieLens Dataset

Predict the name of movies and based upon the reviews of the other critics having similar taste.

The combined dataset consists of 4 different dataset. The links dataset have 3 variables and 9k+ observation, the movies dataset have 3 variables and 9k observation. The ratings dataset have 4 variables and 100k observation. The tags dataset have 4 variables and 1297 observations.



Pokemon Dataset(LDA)

Build a pokemon dream team of 6 pokemon that inflicts the most damage while remaining impervious to any other team of 6 pokemon.

The dataset contains 41 variables and 801 observation of data on pokemon from all 7 generations.



Property Inspection prediction

Predict a transformed count of hazards or pre-existing damages using dataset of property information.



Letter recognition

Predict the letter category based on its attributes.

The test dataset contains 17 variables and 4000 observations. the train dataset contains 18 variables and 16k observation of 26 capital letters of english alphabet based on their different attributes.



DATA ANALYSIS WITH R



Kickstarter funding patterns

Derive insights on successful and failed projects on Kickstarter platform

The dataset contains 15 variables and around 400,000 observations



How bad is the Air Quality in metropolitans?

Analyse worsening airquality in metropolitan cities

The dataset contains 6 variables and over 100 observations



Marketing strategies in Retail banking

Derive insights on how sucessful are the direct marketing campaigns of a Portuguese Bank

The dataset contains 17 variables and over 45000 observations



What are the Characters in superhero comics

Identify the good, bad and the ugly nature of characters in Marvel comics

The dataset contains 11 variables and over 20000 observations



What caused International crisis

Identify all factors which caused major international crisis events in the last 100 years

The dataset contains 96 Variables and over 1000 observations



DATA VISUALIZATION WITH TABLEAU



Hubway data visualization challenge

Produce visualizations that reveal interesting user patterns about how people in Boston gets around on Hubway
The dataset contains 1 million observations on bike usage by residents of Boston