

Chaitanya Sudhakar

Contact: 91 - 9398400010 / E-Mail:chaitanya.chowdary25@gmail.com

Data Scientist Machine Learning, Predictive Analytics, Data Science

PROFILE SUMMARY

Offering More Than 5 years of experience in Information technology and software development field focusing in **Data Science, Predictive Analytics and Machine Learning, Time Series Forecasting, Deep Learning and NLP**

- | | | |
|----------------------------|-------------------------------|-------------------------------|
| - Data Science | - R Programming | - Statistical Modelling |
| - Deep Learning Algorithms | - Python | - Time Series Forecasting |
| - Predictive Modelling | - Machine Learning Algorithms | - Natural Language Processing |

- Develops reusable, maintainable and effective predictive and behavioural models using machine learning techniques.
- Experience in using the machine learning algorithms such as **Linear, Logistic Regression, SVM, Decision Tree, Random Forest, KNN, Naive-Bayes & K-means**.
- Data Scientist familiar with gathering, cleaning and organizing data for use by technical and non-technical personnel. Advanced understanding of statistical, algebraic and other analytical techniques.
- Collaborated with internal stakeholders, identifying and gathering analytical requirements for customer, product and project's needs.
- Developed software applications for statistical modelling and graphic analysis to understand data and drive accurate insight.
- Applies intellectual curiosity and deep analytical thinking to mine large data sets for hidden gems of insight and correlation.
- Actively seeks new methodologies, algorithms, tools and technologies to improve existing models and build new state-of-the-art models.
- Skilled in solid data science background with data engineering applied in industry statistical skill; experience on statistical R, Python and supervised and unsupervised algorithms, machine learning techniques.
- Strong familiarity in working with various statistical concepts such as Hypothesis Testing, t-Test, and Chi - Square Test, ANOVA, Statistical Process Control, Control Charts, Descriptive Statistics and Correlation Techniques.
- Hands on Experience in Implementing Sentiment Analysis, Topic Modelling (with LDA, NMF) and Name Entity Recognition(NER) by using Spacy and NLTK Libraries.

ORGANIZATIONAL EXPERIENCE

May 2017 – Dec 2019: IBM India Private Limited as Junior Data Scientist
Dec 2019 – Jul 2020: SG Analytics India Private Limited as Data Scientist
Nov-2020 – Nov-2021: ZS Associates Private Limited as Data Scientist
Nov-2021 – Apr-2023: Evoke Technologies as Senior Data Scientist

Projects Executed:

Heart Failure Prediction

Main project features: Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Heart failure is a common event caused by CVDs and this dataset contains 12 features that can be used to predict mortality by heart failure. Most cardiovascular diseases can be prevented by addressing behavioural risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies. People with cardiovascular disease or who are at high cardiovascular risk. Due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease needs early detection and management.

Objective: People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help.

Activities performed:

- Exploratory data analysis on Patients Data set.
- Identifying key insights/ patterns in the data.
- Convert Some Categorical variables into Dummy Variables and dealing with Missing Values in the Data set.
- Need to Apply feature engineering techniques to check the outliers and imbalanced data in a data set.
- Perform Some Visualisation plots for identifying the relationship between the dependent and independent variables

- Predictive modelling (Random Forest) to predict the mortality caused by Heart Failure
- Grid search parameter tuning for enhancing random forest model performance.

Warranty Claim Status Prediction

Main project features: Fork Uplifting Manufacturer wants to automate their warranty claims process By Predicting “Claim Status” as Approved/Denied and By Predicting “Total Approval Amount” if “Claim Status” is Approved

Activities performed:

- understanding the business process and features of data and apply few of the filters to the dataset.
- Performing Exploratory Data Analysis on the Dataset and Identifying key insights/ patterns in the data.
- Convert Some Categorical variables into Dummy Variables and also dealing with Missing Values in the Data set.
- Performing some feature engineering techniques to identify the Outliers and handling Imbalanced data.
- Performing Feature Selection Techniques of Feature Importance and Correlation w.r.t to target Variable to identify the Features for the Model.
- Building **Random Forest** Model on Train Data and Performing Predictions on test data
- Performing Hyper Parameter optimization to find out the best Parameters for the model.

Customer Exit Criteria from Bank

Main project features: One of the major problems that every Bank face irrespective of the work sector is the Customer Attrition. Proper strategies and ideas are required to control the Customers exit criteria rate. This framework predicts the Customer attrition and helpful Bank to make decisions and build policies accordingly.

Activities performed:

- Understanding the business process and features of the data.
- Data pre-processing and preparation on the data (**Handling Missing Values and Categorical Features**)
- Treating the Outliers and Handling the Imbalanced Dataset by using imbalanced techniques like over sampling, under sampling.
- Building **Artificial Neural Network(ANN)** Model on Train Data and Perform Predictions on Test Data
- Performed model evaluation on test data - Calculating the confusion matrix and plotting ROC curves
- Apply Hyper Parameter Tuning Techniques

Predicting Whether the Customer will subscribe to Term Deposit

Main project features : Term deposits are a major source of income for a bank. A term deposit is a cash investment held at a financial institution. Your money is invested for an agreed rate of interest over a fixed amount of time, or term. The bank has various outreach plans to sell term deposits to their customers such as email marketing, advertisements, telephonic marketing and digital marketing. Telephonic marketing campaigns still remain one of the most effective way to reach out to people. However, they require huge investment as large call centers are hired to actually execute these campaigns. Hence, it is crucial to identify the customers most likely to convert beforehand so that they can be specifically targeted via call.

You are provided with the client data such as: age of the client, their job type, their marital status, etc. Along with the client data, you are also provided with the information of the call such as the duration of the call, day and month of the call, etc. Given this information, your task is to **predict if the client will subscribe to term deposit.**

Activities performed:

- Understanding the business process and features of the data.
- Identifying key insights/ patterns in the data.
- Convert Some Categorical variables into Dummy Variables and also dealing with Missing Values in the Data set.
- Performing some of the feature engineering techniques to check the outliers and imbalanced data in a data set.
- Standardizing the data and feature selection for Dataset.
- Building **Random Forest (RF)** Model on Train Data and Perform Predictions on Test Data
- Performed model evaluation on test data - Calculating the confusion matrix and Accuracy.
- Grid search parameter tuning for enhancing Decision Tree model performance.

ACADEMIC DETAILS

Bachelor of Technology (Electronics and Communication Engineering) from S.R.K Institute of Technology, Vijayawada (2012-2016)

TECHNICAL SKILLS

Operating System	: Windows
Database	: SQL, Oracle
Language	: R, Python, and SQL

Machine learning Algorithms : Logistic regression, Linear Regression, Multi Regression, KNN, decision tree, random Forest, XgBoosting, Naive Bayes and Support Vector Machines and Deep Learning Algorithms.
Deep Learning Algorithms : ANN. CNN. RNN, LSTM RNN and Auto Encoders
Libraries : Numpy, Pandas, Matplotlib, Seaborn, Keras, Spacy, Tensorflow, NLTK,

PERSONAL DETAILS

Address : Hyderabad
Languages Known : English, Hindi and Telugu.