

Prabhakhar

Email : prabhakharmail@gmail.com

Con : +91-8147939931

PROFESSIONAL SUMMARY

- Business-oriented data scientist with 3.5 years of significant experience in data mining, analysis, statistical analysis and developing solutions, methods for data analysis.
- Had Good Knowledge in Data Modeling in python using advanced libraries like SciPy, NumPy, Pandas, Matplotlib, Sklearn, Seaborn.
- Worked python using Jupyter Notebook , Spyder.
- Hands on experience in Machine translation and text processing.
- Worked on Regression models extensively.
- Worked on various machine learning techniques like classification and clustering.
- Worked on support vector machine and XGboost model.
- Knowledge on Hadoop Architecture and Hadoop Distributed file system, EcoSystem.
- Hands on experience setting, configuring, and programming on a Hadoop platform.

EMPLOYMENT DETAILS

- ANUTA NETWORKS PRIVATE LIMITED as Data Analyst Oct 2014 to till date.

Educational Qualification

- MCA in 2014 from Andhra University, Andhra Pradesh.
- B.S.C Computers 2011 from Sri Venkateswara University, Andhra Pradesh.

Technical Summary

Primary Skills : Data Analytics, Machine learning, Statistics, Predictive analytics.
Programming : Java, Python.
Packages : Pandas, NumPy, SciPy, Matplotlib, Sklearn, Seaborn.
Tools : Anaconda, Jupyter Notebook, Spyder.
Web Development : HTML, CSS, JavaScript.
Database : Oracle, MySQL.
Operating Systems : Windows XP/7, LINUX.

Project-1 : Insurance Fraud Detection

Client : Indian Client

Technologies : Python, Jupyter Notebook, Pandas, Scikit Learn.

Advanced analytics can help insurers identify and reduce fraud-related losses, as well as condense the claims cycle, resulting in improved customer satisfaction. The health insurance claims dataset having 100,000 records in total with 56 attributes is considered for

this study. A major challenge in developing a model for detection of fraudulent health insurance claims was lack of fraud indicator in the data.

Responsibilities:

- Classification of Insurance Claims that are fraudulent and non-fraudulent.
- Analyzed customers data related to their Insurance Payment, Customer Data, Other Insurance Products Subscribed Data.
- Merged tables based on the Insurance-ID attribute and stored the data in a csv format.
- Obtained 3000 Insurance and 80k+ records with 72 feature columns.
- Processed the data using pandas i.e, cleaning, filling missing values and removing outliers.
- Through co-relation found that 15 columns were influencing the target variable i.e, Loan default
- Used classification algorithms such as SVM, logistic regression, XGBoost and performed hyper parameter tuning by gridsearch.
- Have been able to get around 83% accuracy in the model by using XGBoost.

Project-2 : Customer Segmentation across the Bank.

Client : Indian client

Technologies : Python, Jupyter Notebook, Pandas, Scikit Learn

Client is one of the banking industry leader having branches across major cities and other parts of India. Here i need to build the clustering model there by sharing insights to Client.

Responsibilities:

- Performed Customer Segmentation based on Customer Demographics, Transactions.
- Extracted tabled data from Oracle Database using cx_oracle package available in Python.
- Merged tables based on the Customer-ID attribute and stored the data in a csv format.
- Obtained 3000+ unique customers and 200k+ records with 63 feature columns.
- Processed the data using pandas i.e, cleaning, filling missing values and removing outliers.
- Through co-relation found that only 31 columns were unique to each other and removed remaining columns.
- Used K-means and Hierarchical clustering methods to group customers and found that only 6 clusters gives us better clustering by Elbow-joint method.

Project-3 : Click Stream Analysis

Client : Kohls

Technologies : Hadoop, HDFS, PIG, HIVE, Sqoop.

The Business objective of this initiative is to reach out to customers by email who have

- a. Abandoned carts (added items to cart and did not make purchase)
- b. Abandoned Browse (browsed items and did not make a purchase)

Per the industry guidance an effective carts abandon process can recover 18% and send approximately 17500 emails on average.

Responsibilities:

- Loading datasets to HDFS and writing pig scripts.
- Understanding of Datanode, Namenode, job-tracker, secondary Name node, Task Tracker etc.
- Loading data to hive tables and writing queries to process.
- Resolve issues and Bugs fixing.

Project-4 : SRM Infrastructure Operations(SRM-IO)

Client : SRM University.

Technologies : Java, Spring, Hibernate, Oracle, JavaScript, HTML, CSS, Apache-Tomcat.

This Project Basically Divided into Three Huge Modules one is Admin Module ,Staff Module and User Interface Module. Admin Module can be accessed by Higher Authorities where a Unique Identification Number for Admin. In User Interface Module we will develop, maintain and enhance his Accommodation Details after validating College-ID. All rolls are Under Control of Admin module, staff module to allows add or Edit reference code values.

Responsibilities:

- Design front end screens and coding.
- Responsible for Testing and Bug fixing.
- Resolve issues on data visualization screens.
- Involving in Object Relational modules.