

Sibe Sambathkumar

Data Analyst

Chennai, Tamil Nadu
9943750210
sibesambathkumar@gmail.com

SUMMARY

2.7 years of experience in Data Engineering and Data Warehousing with expertise in Data analysis and development of Business requirements, Business processes and System solutions. Proficient knowledge in Python, SQL, and tools like Power BI, Tableau, SSIS and SSAS. Upskilled in machine learning and having expertise in creating predictive and unsupervised models.

- LinkedIn : <http://www.linkedin.com/in/sibes>

EDUCATION

Course	Institution	Percentage	Year
PGP-DSE	Great Lakes Institute of Management	Excel	2023
BE (Manufacturing)	College of Engineering Guindy, Anna University	82 %	2019
12 th Std	Vidyaa Vikas Matric. Hr. Sec. School	95.6 %	2015
10 th Std	Vidyaa Vikas Matric. Hr. Sec. School	97.6 %	2013

KEY SKILLS

- Programming Language:** Python
- Database Language:** T-SQL, MySQL, SSIS, SSAS (Tabular)
- Visualization Tools:** Power BI, Tableau, SSRS
- Cloud :** Azure Data bricks, Azure Synapse
- Microsoft Office:** MS Excel, MS PowerPoint, MS Word
- Libraries:** Pandas, Numpy, Scikit- Learn, Scipy, Seaborn, Matplotlib
- Machine Learning:** Explanatory Data Analysis (EDA), Supervised Learning and Unsupervised Learning, Regression, Classification, Clustering, Decision Tree, Random Forest, Boosting Techniques, Principal Component Analysis (PCA).
- Statistical Methods:** Hypothesis Testing, Chi-Square test, ANOVA test, Z-test, T-test.
- Soft Skills:** Visual thinking, Teamwork, Requirement Analysis, Problem Solving, Planning, Time Management, Root Cause Analysis.

PROFESSIONAL SUMMARY

Senior Systems Engineer, Infosys Ltd | Business Intelligence Analyst

Dec 2019 – June 2022

DELIVERY EXCELLENCE (CPE TRACK) - MAI

Client Name: Microsoft – USA

- Customer Partner Experience (CPE) is mainly concerned with handling millions of surveys that are sent for the Microsoft's business end such as Microsoft Consulting Services, Sales Experience, Customer Experience and Support Services.
- Generating reports using Power Bi, updated spreadsheets and presented results. Optimized BI reporting capabilities for secure server environment.
- Developed Key metrics derived from raw company data to track improvements in organizational efficiency.
- Created workflow diagrams to clearly demonstrate processes and timelines. Optimized system and platform performance capabilities in most efficient, practical way possible.
- Backtracking the data to create Source Target Mapping (STMs), to assist the team in developing their codes during PSDL migration process.

DELIVERY INSIGHTS – MAI

Client Name: Microsoft – USA

- Collected data from wide-ranging sources.
- Designed customized data collection models for specific visualization tasks
- Developed features based on client requirements, acquired data from the source teams, integrated it based on the requirements and visualized into reports

ACADEMIC – DATA SCIENCE PROJECTS

HOST PRICE PREDICTION FOR AIR BNB RENTALS

Business Problem Statement

To build a machine learning model to predict the optimal pricing for the hosts of Bread and Breakfast (BnB) listings with the detailed attributes about the listings.

Approach

Required data cleaning done as a part of EDA. Association of features with target variables studied using visualization graphs and asserted with the statistical test. This involved a detailed analysis on the data and data pre-processing. Feature engineering based on wordcloud is done to obtain more attributes. Random Forest models are built for the optimal calculation of prices. The final models is deployed using Streamlit app.

Skills and Tools

Machine Learning, Robust models, Decision Tree, Random Forest, Tableau, Python, NLP

COST PREDICTION FOR ACQUIRING A CUSTOMER THROUGH MEDIA CAMPAIGN BUSINESS

Business Problem Statement

The objective is to predict the cost spent on customer on acquiring a customer through different media marketing platforms and to gain insights on how the current marketing strategy is efficient to bring in the customer.

Approach

Required data cleaning done as a part of EDA. Association of features with target variables studied using visualization graphs and asserted with the statistical test. Imbalance ML techniques like sampling and weighted loss function implemented to address the class imbalance. Boosting models are built to calculate the price for marketing strategy.

Skills and Tools

Machine Learning, Decision Tree, Random Forest, Boosting techniques, SMOTEN, Python

MACHINE LEARNING BASED DISEASE PREDICTION MODEL

Business Problem Statement

Build the binomial machine learning model to predict the Parkinson disease. The objective of this model is to classify the patients with the Parkinson disease. Perform the required pre- processing steps prior to model building and evaluate the model performance through proper tuning process.

Approach

Multiple datasets merged to form a final dataset. Evaluated the model performance with appropriate measures. Performed all the required graphical and quantitative exploratory data analysis prior to model building. PCA techniques used to reduce multicollinearity between features. Non- linear classification models are built to classify whether the patient has Parkinson or not.

Skills and Tools

Machine Learning, Supervised Learning, Linear and Non-Linear Classification models, Python

CERTIFICATIONS

- Infosys certified Power BI Developer
- Infosys certified SSIS developer
- Microsoft certified: Azure Data Fundamentals (DP-900)