

Suraj Malpani

surajmalpani23@gmail.com | 📞: (682) 331-2509 | <https://www.linkedin.com/in/suraj-malpani> | <https://surajmalpani.github.io/>

SUMMARY: My motive is to derive actionable insights by analyzing diverse datasets; insights, which will be impactful for the betterment of the businesses. I'm a diligent individual with prior professional experience as a Data Scientist and an Analyst.

Related Coursework:

- Applied Machine Learning
- Big Data Analytics
- Applied Econometrics
- Prescriptive Analytics
- Predictive Analytics with SAS

Technical Skills:

- R (tidyverse), Python (pandas, sklearn)
- SQL, NoSQL, BigQuery, MS Access
- Tableau, Power BI, Alteryx
- Microsoft Excel (VBA), PowerPoint
- Hadoop, Spark, MongoDB, UNIX, Git

Certifications:

- 'Google Analytics Individual Qualification'
- 'Machine Learning A-Z' - Udemy
- 'Exploratory Data Analysis' - Datacamp
- 'MongoDB for Developers'
- 'Reporting with R Markdown' - Datacamp

EDUCATION

M.S., Business Analytics, The University of Texas at Dallas - GPA: 3.61

Dec 2019

B.E., Electronics and Telecom Engineering, University of Mumbai - GPA: 3.35

June 2015

BUSINESS EXPERIENCE

Data Science Intern, Daimler, Portland, Oregon

Jan 2019 – Jul 2019

- Streamlined analyses by developing an R Shiny app that gathers and visualizes the data from a daily updated govt website.
- Predicted warranty claims by analyzing telematics fault codes data. Utilized numpy, pandas, scikit-learn to wrangle data.
- Automated the data extraction and statistical algorithmic process using Python and Alteryx app, saving up to 90% of time.
- Utilized machine learning (regression, clustering) and visualizations (Tableau, Power BI) to find patterns and reduce the manual work in assembling the modules on the truck on accurate locations for the custom orders; saved 40+ hours/week.

Teaching Assistant, The University of Texas at Dallas, Dallas, Texas

Aug 2018 – Dec 2018

- Assisted in designing the coursework, evaluated the assignments and quizzes for courses 'Business Analytics with R', 'Database Foundations' and 'Object-oriented Programming with Python', each class consisting of approx. 60 students.

Data Science Intern, Analytics School Online, Remote

May 2018 – July 2018

- Introduced 2 Case-study based courses on Machine learning topics Logistic Regression & Decision trees using R, Python.

Application Development Analyst, Accenture, Mumbai, India

Oct 2015 – May 2017

- Accomplished ad-hoc tasks utilizing Big data and BI tools, Hadoop, Apache Spark, R, Informatica, SQL and Tableau.
- Developed 100+ SQL Queries (Teradata) and 50+ ETL pipelines for acquiring, modeling & transforming the data to create 17 Products Report files; Scheduled cron jobs using bash shell scripts (CLI) for generating automated data reports.
- Automated the process of manual assurance of naming conventions in DataStage ETL data pipelines by developing a 'Code Review' tool in Microsoft Excel using VBA and a UNIX script, effectively saving approx. 40 hours/month.
- Designed dashboards in QlikView for purposes ranging from ERP to presenting KPIs, insights. Debugged ETL jobs, SQL queries and bash/shell scripts among ad-hoc SDLC tasks in a Project where client has database of 90% of UK's population.

ACHIEVEMENTS

Copart Data Challenge (Machine Learning, R, Predictive Analytics)

Nov 2018 – Dec 2018

Selected to present xgboost model which predicts prices of auctioned cars to Copart's executives and Data Science team.

Texas Level R Competition, Winners 2nd Place

Apr 2018 – Apr 2018

Analyzed New York City's crime dataset containing 5.5 million rows, segregated the data based on boroughs and analyzed crime types based on location, time of the day, employment. Visualized this geospatial data using 'ggmap', 'ggplot2' in R.

ACADEMIC PROJECTS

Retail Data Analysis (Marketing, Predictive, Prescriptive, Analytics, SAS)

Sep 2019 – Present

Analyzed brands of peanut butter to provide suggestions to advance the business using the Scanner Data in SAS.

Tweets Classification for Disaster Identification using Spark (NLP, PySpark, MLlib)

Dec 2019 – Present

Marketing Solutions (IBM Cognos, Watson, Python, Customer Segmentation)

Oct 2019 – Present

Twitter Buzz Prediction (Deep Learning, Regression, Tensorflow, Keras, Scikit-learn)

Nov 2018 – Dec 2018

Developed GLM, Ridge, Lasso, SGD, SVM, KNN Regression models to find the best model to predict Twitter Buzz.

Applied Bagging, AdaBoost and Gradient boosting models. Used PCA to reduce features on this data which had 75+ features.

Big Data Analytics (Map/Reduce, Hadoop, Spark, Sqoop, Hive, Impala, HBase, MLlib)

Aug 2018 – Dec 2018

Forecasting Car Sales (Time-series forecasting, ARIMA, R, Tableau)

Jan 2018 – May 2018

LEADERSHIP EXPERIENCE

Member, Business Analytics Leadership Council, The University of Texas at Dallas

May 2018 – Present

- Extracted datasets using BigQuery and REST APIs to Google cloud (GCP) to make it available for students to work upon.