

SHIVANI SOMAN

Email : shivu.soman@gmail.com | Phone : +17326680966 | Website : <https://shivanisoman.github.io>

SUMMARY : Data Scientist with a Master's Degree in Computer Science and 2+ years of work experience in building solutions that translate data into valuable and actionable insights. Experienced in creating machine learning models and ETL frameworks to drive product growth and stakeholder decisions. Team player and an efficient communicator with a growth mindset.

WORK EXPERIENCE

- **Data Analytics Developer / Data Scientist | Barclays Investment Bank** Feb 2019 – Present
 - Analyzed huge amounts of unstructured financial data from multiple sources and created concise, insightful reporting dashboards and visualizations to enable high level Sales and Client Strategy Managers make data-driven decisions.
 - Created a **Hierarchical DBSCAN** based machine learning model for **clustering** similar clients for peer analysis using multiple features including Revenue, Wallet Size, Risk Weighted Assets, Balances and Resource Consumption.
 - Implemented **topic modeling** of millions of user inputted comments using Natural Language Processing and Latent Dirichlet Allocation (**LDA**) to group comments under various categories for easy organization and retrieval.
 - Improved external to internal contact matching for a Unified CRM project by **+300%** compared to previous techniques using Levenshtein Distance based **fuzzy matching** including nickname and phonetic matching.
- **Summer Developer Analyst | Barclays Investment Bank** Jun 2018 – Aug 2018
 - Worked in Risk Analytics on the Client Clearing Hub system performing analysis and **automation** of report generation to greatly reduce user overhead on client trades using Java, SQL and Bash scripting.
- **Graduate Teaching Assistant | University of California, Los Angeles** Apr 2018 - Jun 2018
 - Taught Course LS40 – Statistics for Life Sciences. Conducted 2-hour semiweekly lab sessions in Python to teach undergraduate students different statistical topics like **null hypothesis testing**, p-value, paired test, F-value, chi-squared test, regression, modeling, **statistical inference**, data analysis and visualization.

TECHNICAL SKILLS

- Python, SQL, C/C++, Java, R, Qlik Sense, Tableau, Matplotlib, Flask, JavaScript, Pandas, Scikit Learn, Numpy, MongoDB, TensorFlow, Keras, PyTorch, A/B Testing, Hadoop/Hive, AWS, Apache Spark, MS Office, Git.

EDUCATION

- **Master's in Computer Science | University of California, Los Angeles** Sep 2017 – Dec 2018
Relevant Coursework : Big Data Analytics, Database Management and Statistical Computing, Bioinformatics, Large Scale Data Mining, Learning and Reasoning with Bayesian Networks, Health Analytics **GPA : 3.93**
- **Bachelor's in Computer Engineering | Maharashtra Institute of Technology, Pune** Jun 2013 - May 2017
Relevant Coursework : Data Mining Techniques and Applications, Business Intelligence, Machine Learning, Data Structures and Algorithms, Database Systems , Operating Systems, Computer Networks **Percentage : 71 %**

PROJECTS

- **Music Genre Classification using Spectrograms and MFCC Features** Sep 2018 - Nov 2018
 - Developed a two-fold method to classify music into 10 different genres using Convolutional Recurrent Neural Networks (**CRNN**) for spectrogram analysis and traditional machine learning classifiers for analysis of Mel-Frequency Spectral Coefficients (MFCCs) derived from the audio samples with an accuracy of **86%**.
- **Location Prediction (Big Data Metagenomic Classification)** Apr 2018 - Jun 2018
 - Performed predictive analysis on huge amounts of metagenomic data (**> 3 TB**) to correctly predict the origin of each metagenomic sample using neural networks and **XGBoost**.
- **Activity Monitoring using LSTMS (Health Analytics using SmartWatch)** Mar 2018
 - Built a machine learning algorithm using Long Short-Term Memory Networks (**LSTM**) by collecting data from various sensors like accelerometer and gyroscope obtained from a smart watch to accurately predict the activity the user was performing like standing, sitting, walking, etc.
- **Twitter Popularity Prediction – Super Bowl 2015 Team Sentiment Analysis** Jan 2018 - Apr 2018
 - Analyzed millions of tweets before and during Super Bowl XLIX (2015) using **sentiment analysis** to determine how and why public sentiments towards the New England Patriots and the Seattle Seahawks changed over the span of the championship match.
- **Personalized Medicine : Redefining Cancer Treatment (Kaggle)** Nov 2017
 - Developed a Machine Learning model using NLP that classifies genetic mutations of cancer genes from an expert annotated knowledge base and text-based clinical literature into a set of predefined classes.
 - Obtained a better score than the **1st rank** on the leaderboard of this Kaggle competition using **Word2Vec Embeddings** and **LightGBM**.

EXTRA-CURRICULARS

- Selected to participate in Citadel's data science hackathon – **The Data Open**, where our team analyzed how delays and cancellations in flights affect the revenue and stocks of airline companies.
- Completed Andrew Ng's five-course **Deep Learning Specialization** on Coursera involving Deep Neural Networks, Natural Language Processing, Computer Vision and time series analysis with Sequence Models.