SHIVANI **SOMAN**

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

WORK EXPERIENCE

- **Data Analytics Developer | Sales and CRM Analytics | Barclays Investment Bank** **Feb 2019 – Present**
  - Analyzed high amounts of unstructured financial data and created concise, insightful reporting dashboards and visualizations to help salespeople make data-driven decisions.
  - Incorporated machine learning techniques like forecasting, clustering for peer analysis and topic modelling of user comments and communicated insights and findings to multiple stakeholders.
  - Improved external to internal contact matching by **+300%** using fuzzy matching techniques including nickname and phonetic matching compared to previous SQL-based techniques.
  - Pioneered in creating extensions in QlikSense using Flask, JavaScript and Python to enable fetch/writeback to SQL Server.
- **Summer Developer Analyst | Risk, Finance and Treasury | Barclays Investment Bank** **Jun 2018 – Aug 2018**
  - Worked 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 | Statistics for Life Sciences | UCLA** **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, data analytics and visualization. Also responsible for assignments, midterms and finals grading.

TECHNICAL SKILLS

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

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 (LSTMs) 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) 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.

RESEARCH PAPER

- **Cloud Forensics : Drawbacks in Current Methodologies and Proposed Solution** **Feb 2017**
  - A paper surveying current works in Cloud Forensics and proposing a new model for efficient forensic data and log collection from virtual machines to aid forensic analysis. (IJERA, Volume 7 - Issue 2, Part - III)

EXTRA-CURRICULARS

- Head Organizer at Texephyr 2017, Pune, India for CodeStorm - a vastly popular coding competition.
- 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 involving NLP, CNNs and Sequence Models.