

# Vidhi Gandhi

[Email](#) | 408-309-0620 | [LinkedIn](#) | [GitHub](#) | Milpitas, CA

## EDUCATION

**Santa Clara University, Leavey School of Business**

*Santa Clara, CA*

**Master of Science in Business Analytics**

*December 2020*

**Relevant Coursework:** R, Python, SQL, Machine Learning, Econometrics, Marketing Analytics, NLP, Deep Learning

**MPSTME Mumbai, NMIMS University**

*Mumbai, India*

**B. Tech in Computer Science**

*June 2016*

**Relevant Coursework:** DBMS, Image processing, HCI, Calculus, Data Warehousing and Mining, Big Data

## SKILLS

- **Languages:** Python, R, SQL, SAS, SPARK.
- **Frameworks:** Pandas, NumPy, Scikit-Learn, keras, Matplotlib, Seaborn, NLTK, dplyr, and ggplot2.
- **Data Science:** Statistics, Machine Learning, A/B Testing, Visualization, Marketing Analytics, Data Visualization.
- **Certifications:** [Tableau Data Scientist](#), [SAS® Academy for Data Science Badges](#), [Advanced Google Analytics](#), [AWS Data Analytics Fundamentals](#), [Project Management for Beginners](#)

## EXPERIENCE

**Box**

*Redwood City, CA*

**Practicum Student**

*January 2020 – December 2020*

- Devised churn model for 35000 users to define high probability churn customers and analyze their behavior to improve retention.
- Engineered 14 day cohort analysis to optimize customer retention after onboarding process and understand customer behavior.
- Defined and derived North Star metrics for Product Adoption team, boosting retention for customers with high churn probability by 15%.
- Deployed Random Forest model to predict top 5 features impacting customer churn. These were utilized for A/B testing to target customer by product adoption team.
- Collaborated with stakeholders from 4 teams to identify key problem areas and suggest important feature for user onboarding process.

**Ernst & Young**

*Mumbai, India*

**Analyst, IT Risk Assurance**

*June 2016 – July 2018*

- Performed data extraction and analysis on 3 million rows of audit data from 8 large bank/insurance firms to produce actionable insights and analysis using data visualization for ITGC/SOX audits.
- Formulated automated procedures for assessing Identity and Access Management (IAM) infrastructures for 30,000 employees, reducing weekly efforts by over 75%.
- Performed anomaly detection on client database systems using Python and MySQL
- Received Spotlight award for a large scale project involving 5 global teams. I mentored 4 junior analysts and identified, reported and resolved a crucial error, strengthening the User Access management process.

## ACADEMIC PROJECTS

- **Stock Price Volatility Prediction using CNN, RNN:** Generated Glove embeddings from 17000 SEC 8k filings of all S&P 500 companies. Developed several combinations of CNN, RNN and MLP Architectures, concluded with accuracy of 64% to forecast stock volatility immediately after SEC 8k filings with CNN, RNN combination.
- **Recipe Generator using RNN :** Built a model by training 100k recipes utilizing LSTM and RNN that recommends an entire recipe along with cooking instructions and ingredients when user inputs an ingredient with accuracy of 70%.
- **US Unemployment Rate Time Series Analysis:** Analyzed the monthly U.S. unemployment rate for 900K observations spanning 27 years. Implemented STL decomposition and fit an ARIMA model on R to the time series and forecasted rate for next 4 years.
- **Social Media Analytics:** Prepared, cleaned and reviewed in depth WhatsApp group chat data and real-time Spotify music trends. In Top 100 songs, discovered most similar/different songs and used K-means to cluster similar songs on basis of audio features.