Sabiha Hussain Barlaskar
San Francisco State University
Phone: +1 (512) 975 8391
Email: sabihabarlaskar90@gmail.com | Github: https://github.com/sabiha90 | Linkedin: https://www.linkedin.com/in/sabihabarlaskar90

WORK EXPERIENCE

RAPIDPRICER San Francisco, CA | May - August, 2018

Data Scientist Intern

- Cleaned 2.3 million records of sales data, checking for quality, accuracy and completeness, using Python and performed visualization using d3.js to perform descriptive analysis of the data.
- Loaded the cleaned data into Pandas DataFrame and performed time series forecasting analysis of future sales, using Simple Exponential Smoothing (SES) and AutoRegressive Integrated Moving Average (ARIMA) models.
- Performed multi-class classification of fruit types using Transfer Learning and Convolutional Neural Networks (CNN) with 2.5x times better performance than existing algorithms.

CAPGEMINI INDIA PVT. LTD Bangalore, India | November, 2014 – December, 2015

Software Engineer

- Developed a model using cluster based approach for detecting fraud rings in e-commerce data using Hadoop and Neo4j.
- Performed ETL processes for a telecommunication client to ingest data from SQL relational databases to Hadoop infrastructure using Sqoop.
- Used HIVE for wrangling the data and dispatched it to Teradata to be deployed in production.

PROJECTS

Wrangled OpenStreetMap Data using Python and MongoDB January, 2017

- Transformed OpenStreet map data of the city of Austin, TX into JSON from XML using Python.
- Cleaned and audited the data for quality, correctness and completeness, removing ambiguous zip codes, changing abbreviated street names to full forms and changing inconsistent phone numbers to a standard format.
- Loaded the JSON data into MongoDB and performed exploratory data analysis, such as retrieving total zip codes and finding popular cuisines using MongoDB queries.

Estimated Correlation between Sentiment of Bitcoin tweets with Bitcoin prices January - May, 2018

- Removed missing rows, duplicate rows, unnecessary characters, stopwords and tokenized 41K tweets related to Bitcoins extracted using Tweepy API.
- Loaded the data into a Pandas DataFrame to perform text mining on the data such as creating WordCloud to identify top frequent words and created a frequency matrix using a Bag of Words.
- Performed a sentiment analysis on the data using Maximum Entropy classifier to identify the sentiment of a tweet with an F1-score of 0.92, achieving better results than lexicographic tools like Vader and TextBlob.
- Performed a comparison of the sentiment trend during a period with the trend of Bitcoin prices over the same period to visualize the correlation between them using Tableau.

Built Classification models for Image Processing October - December, 2017

- Reduced dimensions of features using Principal component analysis (PCA) and classified human faces into male and female gender, using Linear Discriminant Analysis (LDA).
- Performed mean shift clustering to segment images using Gaussian kernel function.
- Developed a model to identify human faces by using k-nearest neighbors classification with 85% accuracy.

EDUCATION

- San Francisco State University - *Master of Science in Computer Science* San Francisco,CA | 2017-2019
  - GPA 3.9/4.0
- Data Analyst Nanodegree at Udacity Austin TX | 2016 - 2017

SKILLS

Programming languages: C, C++, Python, R, Java, Perl, Matlab, JavaScript, CSS, PHP, Matlab, HTML
Databases: SQL, Hive, Neo4j(NoSQL), MongoDB, PL/SQL

Frameworks: Scikit learn, Keras, NumPy, Pandas, RStudio, d3.js, Hadoop, Node.js, BootStrap, Tableau, Flask, LUIS, ElasticSearch, Lucene

Cloud: Microsoft Azure, GCP

PUBLICATIONS

- "**Creating Graph Databases From SQL Relational Databases: An Implementation**", published in International Journal of Advanced Research in Computer Science, Volume 5, No. 7, September-October 2014
- "**From Explaining How Random Forest Classifier Predicts Learning of Software Engineering Teamwork to Guidance for Educators**", published in IEEE FIE conference October 2018, San Jose