Sanket Chobe

Chicago, IL, sanketchobe@gmail.com, (725) 266-1351, linkedin.com/in/sanketchobe

Profile

- Data Engineer with 6 years of experience in Software Development for banking and finance domain.
- Proficiency in Python, SQL, Data Analysis and problem-solving using data structure and algorithms.
- Research journal publications in **Machine learning** - Data mining, Social Network Analysis and Clustering.

Work Experience

Data Engineer - Capital One/ Wipro Limited, Chicago, USA: 03/2019 – Present

- Developed Scala, Python, Spark SQL and Snowflake ETL scripts to move Capital One partner data into AWS cloud.
- Automated the process of data preprocessing with Spark Scala and reduced 25% efforts of manual data cleaning.

Software Developer Innovation Intern - Caesars Entertainment, Las Vegas, USA: 05/2017 – 01/2019

- Developed complex **PostgreSQL and T-SQL queries** to visualize the sports betting summary on the leaderboards at The Book - LINQ Casino (Leaderboard as a Service).
- Developed a POC for Quiz, Tone, and Personality analysis chatbots using **IBM Watson** and **Microsoft Azure.**

Analyst - Principal Financial Group, Pune, India: 12/2014 – 07/2016

- Developed COBOL DB2 **SQL** modules for Billing and Specialty Benefits division using Agile SCRUM/XP.

System Engineer - Tata Consultancy Services, Pune, India: 09/2011 – 12/2014

- Developed generic **DB2 SQL Stored Procedures** for New Account Opening application of Morgan Stanley.

Academic Qualification

- University of Nevada Las Vegas, USA, **MS Computer Science Major, GPA - 4.0/4.0**
- Government College of Engineering Amravati, India, **BTech Information Technology, GPA - 8.72/10.0**

Research Publications and Projects

Mining Association Rules for Low Frequency Itemsets - PLOS ONE, 2018. DOI

- Published a **novel association rule mining algorithm** for low frequency and low utility item sets. Achieved same run time efficiency as **FP-growth algorithm**.
- Generated new priority-based association rules to provide effective discount offers or online recommendations.

Advancing Community Detection Using Keyword Attribute Search – Journal of Big Data, 2019. DOI

- Published a **novel clustering algorithm** for **personalized community detection** by creating a new **classification** and **keyword search method** for attributed graphs.
- Data preprocessing – created synthetic graphs by assigning random attributes on nodes of ground-truth network datasets (10,000 nodes) and edges with at least 30% probability of edge creation.
- Achieved a scholar position in the **Data Scientist fellowship challenge** of The Data Incubator.

Correction of Spelling Errors and Sentiment Analysis NLP - Python, NLTK, Scikit-learn, Numpy, and Pandas.

- Implemented **HMM** with 85% accuracy of correction of spelling errors in a document with 44000 characters.
- **Naive Bayes Classifier model** for the sentiment analysis of movie reviews with 95% accuracy of classification of the review as positive or negative.
- **Neural Network**, **Decision Trees**, and **Random Forest** algorithms for movie reviews with around 90% accuracy.

Skills

- **Programing: Python**, **SQL**, **Spark SQL**, **Scala**, C++, Golang, HTML, NodeJS.
- **Algorithms: Community Detection, Association Rule mining, Classification, and Regression.**
- **Data Science**: **Pandas**, NumPy, **Scikit-learn**, Matplotlib, NetworkX.
- **Database**: DB2, PostgreSQL, MS SQL Server, Neo4J, **Snowflake**, Apache Spark, **AWS Cloud**.
- **Data Tools:** Spyder, **Jupyter Notebook**, **Azure ML**, Tableau, Git.