Mohit Chandrabhan Patil

mcpatil@scu.edu | +1(408)-495-1613 | www.linkedin.com/mcpatil | github.com/mohitcpatil

EDUCATION

Santa Clara University, Leavey School of Business — Santa Clara, CA
Master of Science in Information Systems — ***June 2021***
Relevant Coursework: Data Science with Python, Object Oriented Programming Analysis with Java, Cloud Computing Architectures, Big Data Modeling and Analysis, Business Intelligence and Data Warehousing

Ramrao Adik Institute of Technology, University of Mumbai — Mumbai, India
Bachelor's in computer science — ***June 2014***
Relevant Coursework: Data Structures, Analysis of Algorithms, Robotics and Artificial Intelligence, Data Warehouse and Mining

TECHNICAL SKILLS

- **Programming languages:** Python, Java, C++, JavaScript, Scala
- **Databases:** MySQL, PostgreSQL, MongoDB, Apache Cassandra, NoSQL, Hbase, SQL Server, Oracle11g, Redshift
- **Cloud Platforms:** AWS EC2, S3, Google Cloud, Lambda | **Big Data:** MapReduce, Hive, Hbase, HDFS, Pig, SQOOP
- **Data Visualizations:** Tableau, R, Microsoft SQL Server, MS Excel, SSIS, SSRS, Pivot tables
- **Frameworks:** Spark, Hadoop, Cloudera, AWS ElasticMapReduce (EMR), Airflow
- **Libraries:** NumPy, Scikit-learn, Seaborn, PyTorch | **Tools:** Postman, Jupyter, Workbench, CA7, uDeploy, Informatica

EXPERIENCE

Tata Consultancy Services — Pune, India
Application and Data Engineer — 2015-2019

Banking Services (Big data, ETL, Hive, Shell Script, Hadoop HDFS)

- Developed shell Scripts and SQL Scripts for generating reports and running job batches
- Initiated migration of Oracle database to Big Data Hadoop while working in anti-money laundering domain
- Framed the mappings in ETL also modified the ETL code after identifying changes
- Wrote Hive queries and created dashboards for Hive task/Jobs for debugging and fixing
- Migrated Oracle data using SQOOP to HDFS and processed by writing business logic

Retail Domain (Hbase, Cloudera, Python, Hadoop)

- Developed utility scripts in python and shell for purging the data and getting the details of HDFS usage of users
- Setup and installed the Hadoop environment for 8 nodes as well as managed the HDFS disk space to free up
- Analyzed the hive queries and optimized it as per requirement and setup Hadoop clusters across environment in Cloudera

Chatbot Development (AWS, NLP, Alexa, Elastic Search, Lambda)

- Developed Chat-bots on different use cases on platform Alexa skills and Google assistant using Python and Node.js
- Devised a model for Front-end and Back-end application for chat-bot using Angular framework and Typescript
- Implemented third party API's (Weather, Food-Panda, Travel) for developing conversational skills & bots

Quality Kiosk Technologies — Mumbai, India
Business Intelligence Intern — 2014-2015

- Implemented Statistical models also figured customer lifetime value using Predictive modeling technique which include linear and logistic regression, Cluster analysis and other statistical techniques using SQL and R
- Performed Market Basket analysis to identify products which are likely to go together helped in targeted campaign
- Designed BI dashboards to show important metrics/KPI's with the data using insightful drill-down in Tableau

ACADEMIC AND INDIVIDUAL PROJECTS

Music Streaming App Data Pipeline (AWS S3, Redshift, Spark, Airflow)

- Built ETL pipeline to extract log & song data, Spark Data Frames in S3 stage in Redshift & transformed to analytics
- Scheduled, Automated tests & monitored data pipelines using Apache Airflow to ensure Data quality.
- Leveraged Concepts of Distributed System frameworks like MapReduce working with Hadoop HDFS and Hive

Data Analysis on Marathon Runners using Python (Python, Scikit learn, NumPy, Pandas, Seaborn)

- Analyzed the marathon runners split finished times to determine top finishers as per the success ratio of the country
- This analysis helped to find out the strategies used by top finishers (Positive or Negative split) to complete the marathon
- Implemented the models with linear regression and matplotlib which helped to figure out 3 interesting findings

Agriculture Monitor System for managing Agricultural Activities (PostgreSQL RDS, Apache Cassandra, AWS Redshift)

- Created data model of user activities for database and ETL pipeline in Cassandra by defining the Fact and Dimension table
- Performed analysis on user activities by creating the analytics tables from those staging tables created on Amazon Redshift