GAURESH CHAVAN

Data Enthusiast | Information Systems Graduate Student | Former Data Engineer Co-op

(857) 269-9394 https://www.linkedin.com/gaureshc/ chavan.ga@husky.neu.edu karmickoala31

EDUCATION

Master of Science in Information Systems **December 2019**
Northeastern University, Boston, MA

Bachelor of Engineering in Information Technology **May 2017**
Vidyalankar Institute of Technology, Mumbai, India

TECHNICAL SKILLS

Programming : Java, Python, SQL, Painless Scripting (Lucene Query), MapReduce, Hadoop
Database : SQL Server 2017, MySQL, PostGRE SQL, Oracle 11gR2, MongoDB, CosmosDB
Visualization : Tableau, Power BI, Qlik Sense
ETL : Talend, SSIS
Cloud service : Microsoft Azure

WORK EXPERIENCE

Data Engineer Co-op **January 2019 – June 2019**
Novant Health Inc., Charlotte, NC
- Defined, configured and optimized various MDM processes and stages including data acquisition, standardization, harmonization and data provisioning using Talend Data Integration tool
- Automated data ingestion pipeline to minimize data latency for service line applications by 30%
- Led efforts to analyze raw data for source/target mappings, created SQL stored procedures to apply business rules on it
- Accomplished a full text search using Elastic Stack featuring synonym dictionary and fuzzy matching to simplify comprehension of patient clinical notes for physicians
- Worked effectively in a start-up like environment as part of a research oriented, agile, fast-paced delivery and sustainment team

Key Achievement: Recognized across teams and divisions for standing up MDM strategy and expediting data retrieval for production applications by 50%

ACADEMIC PROJECTS

Contoso Retail Data Warehousing **February 2018 – April 2018**
- Translated business requirements into a star schema Data Warehouse using Toad Data Modeler
- Built data migrations jobs in Talend Data Fabric to populate Facts and dimensions from 4 separate data sources
- Implemented an automated master/root job to achieve parallelization and migrate data synchronously with minimum latency (10 million records in 7 minutes approx.)
- Prepared a dynamic and interactive visualization dashboard with Qlik Sense to present patterns in data pertaining to sales and revenue

Implementing Machine Learning algorithms for Cancer Prognosis **February 2018 – April 2018**
- Analyzed Wisconsin Prognostic & Diagnostic Breast Cancer (WPBC & WDBC) datasets from UCI Machine Learning Repository to classify tumor type and predict cancer recurrence & survivability respectively
- Implemented Recursive Feature Engineering to automate feature selection and ran TPOT classifier to obtain optimal pipeline for a prediction problem
- Performed optimization using Auto Scaling and Grid Search techniques to improve overall prediction accuracy from 88% to 92%
- Presented results in form of a portfolio and a research paper (IEEE standard)