

REEMA YADAV

Boston, MA | (857)269-9106 | yadav.ree@northeastern.edu | [LinkedIn](#) | [GitHub](#)

EDUCATION

Northeastern University, Boston, MA

Expected Dec 2021

M.S. in Data Analytics Engineering

Relevant Courses: Database Management & Design, Algorithms, Data Warehousing & Business Intelligence, Data Visualization

University of Mumbai, India

Jun 2016

B.E. in Electronics

PROFESSIONAL EXPERIENCE

SMS Group Inc, Pittsburgh, USA | Data Engineer

Jan 2021 – Aug 2021

- SSAB Asset Health – Utilized python to implement supervised machine learning technique for time series signal anomaly detection on **4 year** unstructured data to predict future failure for furnace and burner block - reduced the downtime by **40%** of the furnace. Increased the number of heats from **600 to 1000**
- KAPH Framework – Fabricated an end-to-end framework for time series data analysis using Python libraries to detect anomalies in steel plant. ETL via Pentaho, S3, Oracle SQL server
- SMS Data Factory - Real-time data integration, pipeline monitoring and data visualization capabilities delivered in a single, and easy-to-use platform

AlSkunkworks, Northeastern University | Research Assistant

Jul 2020 – Jan 2021

- Worked on design of interpretable machine learning models like Shapely and Lime
- Conducted hands-on-experience workshops on exploratory data analysis and SQL for **80+** students using Python

TATA Consultancy Services, Mumbai, India | Data Engineer

Dec 2016 – Dec 2019

- Modeled PoC for BIRD Hadoop Project. Migrated historical customer transactional data of **10 years (700 TB)** in Hadoop and compressed data by **79%** using Snappy compression and Hive configuration
- Responsible for the design, development, testing, documentation and analysis of features for Hadoop project
- Developed design approaches to validate findings. Automated the daily data-ETL process reducing downtime by **82%**
- Deployed MapR Hadoop clusters and did related developments on Production and Disaster Recovery (DR) environment. Administered OS, Database and MapR Application upgradation on **27** Linux servers
- Designed and developed the Big Data Intelligence Reporting Department portal in Java for Bilingual language to extract data from Hadoop clusters for inquiry reducing time for process from **1-2 days to 5-10 minutes**
 - Built Grafana dashboards by ensuring data integrity, analyzed and presented insights to the client which improved customer satisfaction by **30%**

TECHNICAL SKILLS

Tools: Tableau, SQL (Oracle, MySQL, MS), Visual Studio, Jupyter, S3, Pentaho, Azure DevOps, Git, SSIS, AWS, CI/CD

Languages: C, C++, Python, R, Shell scripting, SQL, Hive, Java

Big Data Technologies: Hadoop, Sqoop, Spark, Hive, HDFS, Data modeling, ETL, Data Lakes/ Warehousing

Python/R: Matplotlib, NumPy, Seaborn, Pandas, Scikit-Learn, Bokeh, dplyr, tidyverse, ggplot2, tidy

ACADEMIC PROJECTS

Analysis of Sales via Warehouse Management

Spring 2021

Performed transformation via slowly changing dimension, merge join, sort, derived columns, conditional split, data conversion to get dimension and fact table by SSIS. Data loaded into a warehouse through ETL. Designed OLAP cube

Predicting Buying Intention of Bank Customer

Fall 2020

Created classification model to determine whether the customer will buy a term deposit plan using algorithms- KNN, Decision Tree. Evaluated performance of models - ROC curve and achieved **89.99%** accuracy on the logistic regression

Graduate Student Government at Northeastern University

Summer 2020

Dashboards for finance department with Python and Tableau, cleaned and manipulated data with pandas function. Visualized data through line chart and pie charts with matplotlib. Improved the budget for 2020-21 by **68%**

Supplier and Inventory management Database Design

Spring 2021

Cultivated a RDBMS for gaming e-commerce website with MySQL server. Designed database to Boyce- Codd normal form to curb insert, update, delete anomalies and avoid data redundancy. Implemented new order auditing via triggers, stored procedure to track inventory

LEADERSHIP/ACHIEVEMENTS

Vice President of Academic Affairs | Northeastern University

Data for Good Hackathon Winner 2021 | JPMC