Abhishek Patil
Boston, MA * 617.516.3798 * patil.abhis@northeastern.edu * LinkedIn

Education

Northeastern University, Boston, MA   Pursuing
Master of Science in Information Systems, **GPA 3.7/4**
Solapur University, Maharashtra, India.   March 2020
Bachelor of Engineering in Computer Science, **GPA 3.5/4**

Technical Skills

- **Big Data Technologies** - Hadoop, MapReduce, Hive, Pig, NoSQL, Postgres SQL, Spark, Web Scrapping (Auto Scraper), Informatica, BRIO, DBT, Data Warehousing, Unix
- **Apache Spark** – Spark Core, Data frames, Spark SQL
- **Programming Languages & Technologies**– C, C++, Python, R, Minitab, Java, Java Swing, MongoDB, Scala, Angular, NodeJS, MATLAB, Firmware APIs
- **Cloud Platform (Azure, AWS, GCP, Snowflake)** – Azure Data Factory, Databricks, Data Lake, Airflow, Cosmos DB, GCP, Synapse, AWS Redshift, Azure DevOps, Apache Kafka, GCP Looker Studio, Big Query, Quick Insights, Kubernetes, Docker
- **Data Visualization** – Tableau, Power BI, Quick Insights, Looker Studio, Snowflake Visualization
- **Tools/Software** – Git, Spring Boot, Maven, Talend, NetBeans IDE, Visual Studio Code, JIRA, Junit, SPSS, Terraform, SAS
- **Other Skills** – Asset Liability Management, Statics, and process improvement, data modeling, Microsoft Excel, Microsoft PowerPoint, Docupedia management, Agile Project Management, Risk Management, Business Intelligence and analysis
- **Library** – Pandas, NumPy, SciPy, Matplotlib, SciKit-Learn, Scrapy, Beautiful Soup, TensorFlow

Professional Experience (Key Deliverable)

Data Research Engineer Intern | Pison Technology | Boston | USA   Sep 2023 – Dec 2023
- Collected EEG and EMG data, validated using Python and plotted time series, power frequency to study neuromuscular functioning
- Researched medical response time delay on **Big Query** at different alcohol doping levels and represented on **Looker Studio dashboard**.
- Executed Gesture Recognition using various **Machine learning algorithms** with an accuracy of **86%** and used **QLIK** for analytics
- Enhanced neuromuscular data analysis by leveraging **Snowflake** for efficient data aggregation , integration, real-time visualizations

Big Data Engineer | Infosys Ltd | India   Nov 2020 – June 2022
- **E-Bill Generation and Payment Service -** Technically **Analyzed and gathered data** from various sources, generated detailed e-bills for **18.8 million customers** using **spark data frames** notified them via email, and added flexibility to pay later
- Incorporated concepts of **partitioning, indexing, and bucketing** for faster retrieval of Financial reports through data analysis of sales.
- Increased the efficiency of the **Azure Data Factory** pipelines by up to **30%** by reducing processing time and cost charged by the cloud service providers by using **advanced SQL** queries like PIVOT, Window functions, multi-nested queries
- Led the team in constructing **data architecture pipelines** that are **scalable, repeatable, and serve multiple purposes**
- Automated **ETL** procedures, reducing the time by up to 40% by creating predefined **PL/ SQL** scripts and using **Airflow** for orchestration

Athletics Operations Assistant Data | Northeastern University | Boston | USA   June 2023 - Present
- Designed and implemented a dimensional data model with SCDs in Amazon **Redshift**, boosting insight speed by 30% query performance by 40% with **Amazon RDS** , **DynamoDB** and Stored in **S3** buckets
- Streamlined Financial audit data transformation via **AWS Glue** ETL, enhancing data reliability for analytics, reducing processing times by **40%**, and supporting advanced post match analysis in **Amazon Redshift**
- Led the development of a secure data lake with AWS Lake Formation , facilitating real-time analytics with **Amazon Kinesis** and dynamic visualizations in **Amazon Quick Sight**, improving game-time decision-making

Data Analytics Intern | Aim Script Technologies Ltd | India   Dec 2018 – Jan 2019
- Designed dynamically responsive dashboards utilizing **Tableau** to summarize Data Quality Metrics of Health disease data.
- Proficient in **dimensional data modeling** with ER/Studio, Erwin, Sybase Power Designer, **FACT & Dimensions tables**, Star Join Schema/Snowflake modeling, Relational Data modeling, and Conceptual, Physical, and Logical data modeling.
- Using **MS Excel's, Google Sheets** pivot and lookup functions, I tracked business metrics and produced reports on demand

Academics

We Care Scrutiny | Northeastern University
- Demonstrated use of **Java Swing, SQL, Power BI, Azure (Databricks, Data Factory, cosmos DB)** for dynamically comparing the insurance policies of various companies
- Achieved Statistical representation of data using **Power BI** regarding the **most popular** insurance plans with an accuracy of **86%** and also gave an analysis of customer traffic to insurance companies so that they can design new plans accordingly
- Developed **stored procedures, triggers, views, and functions** in SQL server to optimize data extract, transform, and load(**ETL**) processes by **30%**

Better Purchase Online Shopping Indicator | Northeastern University
- Established **Data Cleaning and Pre-processing** using **pyspark** in **Azure Data Factory** in order to load data from files to staging databases with appropriate datatypes.
- Suggested the most optimal website for purchasing a product by performing transformations using SQL queries to compare over various factor's like price, expiry_date, quantity, brands, quality ,variety and loaded the tables to **integration** database
- Leveraged **Azure Cosmos DB** for scalable data ingestion and employed **Azure Synapse** Analytics for advanced analytical transformations to facilitate optimal product purchasing decisions.