BHAWANA AGARWAL

agarwal.bh92@gmail.com | (857) 250-6014 | LinkedIn - bhawana-agarwal-neu| GitHub - bhawanaagarwal| Medium - @agarwal.bh

EDUCATION

Master of Science in Information Systems | Northeastern University | Boston, MA **Jan 2022 - Dec 2023**

Relevant Courses: Data Science Engineering Methods & Tools, Database mgt. & design, Data Structure & algorithms, Advances in Data Science, Machine Learning in FinTech, Neural Network modeling

Bachelors in Electronics Engineering | Rajiv Gandhi Proudyogiki Vishwavidyalaya | MP, India **Aug 2011 - May 2015**

TECHNICAL SKILLS

Programming Language:	Python, JavaScript, R
Python Libraries:	Pandas, NumPy, Matplotlib, Sci-kit learn, Seaborn, OpenCV, TensorFlow, Keras
Databases:	Oracle SQL, MySQL, PostgreSQL, MongoDB
BI and Visualizations Tools:	Tableau, Power BI, Alteryx
Machine Learning Algorithms:	Linear Regression, Logistic Regression, Random Forest, Decision Tree, Gradient Boost, K-means Clustering, SVM
Deep Learning:	Neural Network (CNN, RNN, ANN, LSTM), LLMs
Hard Skills:	Data Analysis, Model Evaluation, Hyperparameter tuning, Statistics, Advanced Analytics, Predictive models, SDLC (Software Development Life Cycle)
Cloud Platform:	AWS (IAM, EC2, AWS Sagemaker), AWS Solution Architect Associate (In training), GCP

PROFESSIONAL EXPERIENCE

Data Scientist | Ernst & Young | Bangalore, India **Oct 2019 – Dec 2021**

- Identified key business questions to transform functional requirements into technical requirements and collected large dataset
- Extracted required information from unstructured dataset, preprocessed it (transformed) and packaged it into executable file to load data to Excel saving **$1.2M**
- Collaborated with cross-functional teams to enhance tax preparation process by optimizing complex data extraction, reducing form completion time by **30%,** and improving data accuracy by 20%
- Applied unsupervised learning to segment database of nearly **70k customers**, identifying patterns of high-spenders, savers, and other segments based on variables like income, age, and spending score
- Conducted exploratory data analysis and feature selection on employee data in the tax domain, achieving a 65% accuracy rate in predicting employee attrition with Logistic Regression
- Spearheaded development and promotion of a scalable VAT reconciliation tool for raw data, achieving an **efficiency gain of 95%**

Data Analyst | Tech Mahindra Ltd. | Noida, India **Feb 2017 – Aug 2019**

- Designed Oracle database to manage client engagement data with **15% improvement in data accessibility**
- Visualized key metrics such as response time, resolution time, and feedback ratings using Tableau dashboard for stakeholders, resulting in **20% increase** in identifying support ticket trends
- Enhanced web application performance by replacing complex subqueries with common table expressions CTEs resulting in a **15% reduction in query response time**
- Contributed to the complete SDLC from requirement gathering, design, development, testing, Production, user training and support for production environment

Software Engineer Intern | Zenrays Technologies | Bangalore, India **Jan 2016 – Nov 2016**

- Debugged JavaScript code for web applications, identifying and resolving **25+ issues**
- Created user guides, contributing to a 30% reduction in user queries and facilitating seamless user interaction

PROJECTS

Precipitation Prediction with Geospatial Data | Python, TensorFlow, ConvLSTM, LSTM **Oct 2023 – Dec 2023**

- Brainstormed machine learning (ML) model for precipitation prediction for image and temporal dataset
- Successfully addressed challenges related to imbalanced temporal data, achieving an impressive 80% accuracy in rain prediction

Reddit Data Pipeline | Python, Apache Airflow, Celery, ETL, Amazon S3, AWS Glue, Amazon Redshift **Aug 2023 – Sept 2023**

- Extracted insights from Reddit using its API, employing ETL workflows with Apache Airflow and Celery
- Stored data on Amazon S3, leveraged AWS Glue for data cataloging and loaded data into Amazon RedShift for analytics by setting up Redshift Cluster

Multiclass Image Classification | TensorFlow, Convolutional Neural Network, Python **Nov 2022 – Dec 2022**

- Developed an image classifier attaining a consistent accuracy of 73% across various numbers of image classes