BHAWANA AGARWAL

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

SUMMARY

Analytically minded self-starter with 5 years of experience and familiarity in gathering, cleaning, and organizing data for use by technical and non-technical personnel. Possesses an advanced understanding of statistical and other analytical techniques. Enjoys automating mundane tasks using Python. Has extensive hands-on experience in Python and SQL with expertise in Machine Learning algorithms. Adept at collaborating with cross-functional teams to effectively address business challenges.

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, MS SQL
BI and Visualizations Tools:	Tableau, Power BI, Alteryx, Microsoft Excel
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,
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 **100+ JSON and XML files**
- Developed a Python automation tool, achieving **$1.2M cost savings** by extracting, preprocessing, and packaging key information from JSON and XML datasets into an executable file for loading into Excel
- Collaborated with cross-functional teams to optimize tax preparation, achieving a **30% reduction** in form completion time through enhanced data extraction and process optimization
- Applied unsupervised learning to segment database of **70k customers**, identifying patterns among high-spenders, savers, and other segments based on various features such as income
- Designed machine learning model to predict employee attrition using Logistic Regression with **65% accuracy**
- 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 and implemented Oracle database to manage client engagement data, enhancing data accessibility by 15%
- Created Tableau dashboards for stakeholders, visualizing KPIs leading to 20% increase in identifying support ticket trends
- Optimized web application performance with **15% reduction** in query response time by replacing complex subqueries with common table expressions (CTEs)
- Contributed to complete Software Development Life Cycle (SDLC) using **JIRA** reducing development cycle time by 20%

PROJECTS

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

- Brainstormed challenges related to imbalanced temporal data for precipitation prediction model
- Successfully designed ensemble machine learning model , achieving 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