BINAY CHANDRA

/binaychandra | /binaychandra | binay.ch

Gurgaon, India | +(91) 727-640-9096 | binaychandra90@gmail.com

PROFILE SUMMARY

Experienced Data Scientist with 10+ years in machine learning, deep learning, big data, and Generative AI. Skilled in cloud technologies and agile practices, with a passion for solving complex problems and delivering scalable, data-driven solutions. Strong interest in research, programming, and innovation to drive efficiency and impact.

EXPERIENCE

Senior Data Scientist, Ernst & Young, *Gurgaon, India* **September 2022 - Present**

- **Demand Forecasting (Mondelēz)** - Designed and deployed a cost-efficient demand forecasting pipeline on Databricks and GCP using Random Forest algorithm, achieving ~90% MAPE across 28 parallel models. Demo
- Implemented **MLOps pipeline with MLflow** for experiment tracking, GitHub Actions for **CI/CD**, and **SonarQube** for code quality enforcement; also developed a dashboard to monitor model and data quality, enabling automated, high-quality model delivery.
- **Architected and led** Azure-based Labor Economics solution with **Databricks pipelines**, creating an end-to-end automated framework with **PowerBI** dashboard for data-driven decision making that **improved workforce insights by 35%**
- Developed a **Flask-based data extraction** and quality assessment framework enabling stakeholders to explore data via interactive plots and receive real-time query responses and visualizations powered by **Generative AI**. Demo
- Built a **RAG-based chatbot** for real-time supply and demand information retrieval by integrating multiple SQL databases and enabling fast graph generation and dynamic insights. Demo

Azure | Databricks | PySpark | PyTorch | MLflow | Generative AI | HR Analytics | Data Governance | GCP | Azure DevOps | CI/CD

Data Scientist, Publicis Sapient, *Bengaluru, India* **July 2018 - September 2022**

- Spearheaded the development of **Spark-powered processing engine** for e-commerce data, implementing **image processing** and **NLP algorithms** for Keyword extraction and sentiment analysis that delivered real-time insights to 30+ marketing teams
- **Improved forecasted booking accuracy by 20%** using an ensemble of **Facebook Prophet and BigQuery ML**, with COVID-19 **impact analysis** and automated email notification for Statistical and forecast reports. Demo
- **Media Mix Modeling** Developed an MMM model to **predict and optimize revenue** across marketing channels using campaign data, factoring in spend, seasonality, and external variables; enabled **data-driven budget allocation** to maximize ROI.
- Built automated data pipelines integrating **Google BigQuery, Google Sheets, and marketing APIs**, along with web scrapers for user review extraction to enable **topic modeling, and performance dashboarding** in Google Data Studio.
- **Pioneered a Django-based** voice assistant system for Automobiles that offered seamless control over news, music, and food orders. **Leveraged Dialogflow** and other APIs to understand user intent and deliver personalized experiences. Demo
- **Deep Ads** – Built the creative analysis framework to extract ad features using Google Cloud Vision API, mapped them to marketing KPIs, and developed a predictive model to assess performance and provide actionable recommendations.

PySpark | GCP | NLP | Machine Learning | Cloud Composer | Dataproc | PostgreSQL | Google BigQuery | Time Series | Alteryx | Google Data Studio

Data Analyst, Tata Consultancy Services, *Pune, India* **October 2011 - July 2018**

- **Built a text classification model** to categorize user narratives for **Fiat Chrysler** using **Python**, **NLTK**, and deep learning techniques; enhanced model accuracy through advanced pre-processing and neural network architectures.
- **Designed a recommender system** using **collaborative filtering** to suggest triggers for an automotive diagnostic tool; improved tool efficiency and user throughput significantly using **Python**, **SQL**, and **Sklearn**.
- **Developed a predictive maintenance model** to forecast component failures using telemetry data; presented results via interactive **Tableau dashboards** for stakeholder insights. Tools used: **Python**, **Sklearn**, **SQL**, **Tableau**.
- **Engineered a high-speed data processing engine** using **MATLAB MapReduce** for mining vehicle diagnostics data, reducing execution time by **95%**; integrated with **.NET** and **SQL** systems.
- **Led middleware architecture design** for SOA-based messaging infrastructure at **Morgan Stanley**; connected **.NET** and **Java** services to mainframe backends using **IBM DataPower**, **XSLT**, and **Linux**.

Python | Machine Learning | NLP | NLTK | Sklearn | SQL | Tableau | MATLAB | MapReduce | .NET | IBM DataPower | XSLT | Linux

EDUCATION

M.Tech in VLSI and Embedded Systems, *Pune University, India (Percentage: 70.4%)* **2014 - 2016**
Thesis: "Segmentation based feature extraction of MRI images using Wavelet and its implementation on FPGA"

B.Tech in Electronics and Telecommunication, *Bharati Vidyapeeth University, Pune, India (GPA: 8.3/10)* **2007 - 2011**
Relevant Coursework: Digital Signal Processing, Computer Architecture, Embedded Systems

SKILLS & INTERESTS

Programming & Data Science: Python, MATLAB, SQL, Pandas, NumPy, Scikit-learn, Langchain, PyTorch, Deep Learning, NLP, Recommender Systems, Generative AI, Forecasting Models, Large Language Models

Big Data & Cloud Platforms: Apache Spark, Apache Hadoop, Google Cloud Platform (GCP), BigQuery, Azure, Databricks

Tools & Visualization: Git, Docker, MLflow, Power BI, Tableau, Google Data Studio, Alteryx, Flask, Streamlit, Gradio

Methodologies & Soft Skills: Agile, SCRUM, Analytical Thinking, Problem Solving, Communication, Cross-functional Collaboration, Technical Leadership

PUBLICATIONS

- **Segmentation based feature extraction of MRI images using Wavelet and implementation on FPGA**
  2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), IEEE
  Authors: **Binay Chandra**, Manish Sharma | DOI: 10.1109/ICACDOT.2016.7877566
- **Texture Feature Extraction Methods and Wavelet Standpoint**
  International Journal of Innovative Research in Computer and Communication Engineering, 2016
  Author: **Binay Chandra** | DOI: IJIRCCE Publication

LEARNING & CERTIFICATIONS

- Generative AI with LLMs, *DeepLearning.ai* (2024)
- Scaled Agile Accredited SAFe® 5 Certified, *Scaled Agile, Inc.* (2022)
- Apache Spark 3 - Spark Programming in Python, *Udemy* (2021)
- Apache Airflow | A Real-Time & Hands-on Course on Airflow, *Udemy* (2020)
- Deep Learning with TensorFlow, *Coursera* (2019)
- Machine Learning, Stanford University, *Coursera* (2017)
- Six Sigma Yellow Belt (2016)
- ITILv3 Foundation Certification (2014)

HONORS AND ACTIVITIES

- "Bravo Award" for excellent performance and end to end delivery of the project (2024)
- Awarded ”Star of the Sprint” for successful delivery and optimizations (2022)
- Howathon-Runners up: Developed voice assistant for services like news, games, songs, food order etc (2020)
- Expo-Runners up: Improved Marketing campaign performance with Computer Vision & Machine Learning (2019)
- ”Inclusive Collaboration Award” for creating ideas and enabling others to leverage the same (2019)
- Awarded "Best Team Player" for outstanding performance in the team (2018)
- Awarded "Best Outgoing Student" in B.Tech (2011)
- First prize in Newton’s Apple, Bharatiyam, National level Techfest (2009)

REFERENCES

Sumil Mehta
Sr. Marketing Specialist, FRACTAL.AI
sumilmehta007@gmail.com

Ashish Yadav
Full Stack Developer, WELLS FARGO
ashishkumar150190@gmail.com