**SUMIT WATKAR**

• Pune, India • watkar.sumitratnakar@gmail.com • (+91) 9960862867 • [portfolio](https://github.com/sumitwatkar/Data-Science-Portfolio)

**PROFILE SUMMARY**

T-shaped professional, proficient in understanding customer business problems, deriving insights from the data and identifying the right models. Skilled in **MLOPS** for seamless model integration, deployment, and driving impactful business growth

**PROFESSIONAL EXPERIENCE**

Incedo Technology Solutions Ltd. - Sr. Data ScientistSep 2024 - Present

Capgemini Technology Services India Ltd. - Data ScientistDec 2022 - Sep 2024

Infosys BPM Ltd. – Technology AnalystJul 2022 - Dec 2022

Intech Software Solutions - Data AnalystAug 2020 - Jul 2022

Tecsolvent Software Technologies – Data AnalystSep 2017 - Aug 2020

ZF Steering Gear India Ltd. - Trainee EngineerJun 2015 - Aug 2016

**KEY ROLES AND RESPOSIBILTIES**

* Interact with customers, develop use-cases on ML/AI, and use the same in solution integration on existing business
* Selecting features, building, and optimizing classifiers using machine learning techniques
* Data mining using state-of-the-art methods and handling large-scale data using ETL: processing tools such as **PySpark** and **Databricks**
* Extending company's data with third party sources of information when needed
* Enhancing data collection procedures to include information that is relevant for building analytic systems
* Processing, cleansing, and verifying the integrity of data used for analysis
* Doing ad-hoc analysis and presenting results in a transparent manner
* Creating automated anomaly detection systems and constant tracking of its performance
* Research and devise innovative statistical models for data analysis
* Demonstrated technical proficiency in data manipulation, statistical modeling, and machine learning to solve real-world business problems effectively
* Developed and refined algorithms, sculpting them into powerful tools to overcome challenges
* Conducted in-depth quantitative analyses, navigating vast datasets to extract meaningful insights
* Work as the lead data strategist, identifying and integrating new datasets that can be leveraged through our product capabilities and work closely with the engineering team to strategize and execute the development of data products
* Collaborate with product design and engineering to develop an understanding of needs and Quality
* Analyze data for trends and patterns and interpret data with a clear objective in mind
* Provide deep learning and machine learning-based solutions to Solve various real-world problems

**PROJECTS**

**CREDIT RISK MODELLING**

* Developed a credit risk modeling system using **XGBoost** **Classifier** to optimize the bank's lending strategy, resulting in a 20% reduction in non-performing loans and a 15% increase in loan approval
* Built and integrated Probability of Default **(PD)**, Loss Given Default **(LGD)**, and Exposure at Default **(EAD)** models, enabling accurate credit risk assessment and reducing potential losses by 25%
* Enhanced credit approval processes by leveraging machine learning to improve predictive accuracy, reduce Expected Credit Losses (ECL), and support data-driven decision-making

**RETAIL CUSTOMER ANALYTICS**

* Developed a customer segmentation model using **K-Means** and **RFM** **analysis** to quantify engagement and buying patterns. Integrated **Power BI** to visualize and analyze the segmentation results, leading to actionable insights for business strategy. Achieved a 25% increase in targeted marketing campaign effectiveness and a 20% boost in customer retention rates

**SENTIMENT ANALYSIS**

* Developed and deployed XGBoost model with **Amazon** **SageMaker**. Enhanced accuracy through feature engineering and fine-tuning. Created a user-friendly **Flask** interface for real-time predictions. Implemented CI/CD with **GitHub Actions** and utilized **Docker** for containerization, ensuring efficient, reliable deployments on **AWS EC2** instances

**SKILLS**

* **Programming**: Python, PySpark
* **Databases:** MySQL, PostgreSQL, MongoDB, Oracle
* **Data Visualization:** Excel, Power BI, Tableau
* **Exploratory Data Analysis:** Pandas, NumPy, Matplotlib, Seaborn, Plotly
* **Statistics:** Descriptive & Inferential Statistics
* **Supervised Machine Learning:** Linear Regression, Logistic Regression, Decision Trees, Random Forest, Naïve Bayes, SVM, Gradient Boosting, XGBoost
* **Unsupervised Machine Learning:** Anomaly Detection, K-Means, Hierarchical Clustering, DBSCAN, Gaussian Mixture Model, PCA, t-SNE, LDA
* **MLOPS:** CICD Pipelines, DVC, Docker, Flask, FastAPI, GitHub Actions, MLFlow, Prometheus, Kubernates
* **Big Data:** Hadoop, Spark, Databricks
* **Cloud Computing:** AWS, Azure, GCP

**EDUCATION**

**Bachelor of Engineering -** Nagpur University

**ACHIEVEMENT**

* Coursera Certified Specialization in Machine Learning, and Generative AI with Large Language Models
* Conducted interviews, optimized evaluation processes, and collaborated on recruitment strategies to build a high-performing data science team