**SUDHIR GOSWAMI  
DATA SCIENTIST | AI ENGINEER | ML ENGINEER  
Gurgaon India |** [**sudhir.ai.ml@gmail.com**](mailto:sudhir.ai.ml@gmail.com) **|+91 7024137907  
LinkedIn :** [**https://www.linkedin.com/in/sudhiraiml/**](https://www.linkedin.com/in/sudhiraiml/) **| GitHub : https://github.com/sudhiraiml**

**Professional Summary**

* Data Scientist with 6+ years of experience, starting as a Data Analyst and progressing to developing and deploying real-world AI/ML solutions.
* Proficient in the full data science lifecycle: data collection, cleaning, model building, evaluation, deployment, and monitoring.
* Experienced in machine learning techniques such as regression, classification, clustering, time series forecasting, and recommendation systems.
* Strong programming skills in Python with expertise in libraries like pandas, NumPy, scikit-learn, XGBoost, LightGBM, TensorFlow, and Keras.
* Successfully solved real-world problems including customer churn prediction, fraud detection, demand forecasting, and NLP tasks (e.g., text classification, sentiment analysis, topic modeling).
* Built and maintained end-to-end ML pipelines with real-time deployment and performance monitoring.
* Solid foundation in statistics and A/B testing, ensuring model accuracy, reliability, and interpretability.
* Comfortable working with cloud platforms like AWS, Azure, and GCP for scalable data processing and machine learning workflows.

**Skills**

**Programming & Scripting:**  
Python, SQL, Bash, Git & GitHub, Jupyter Notebook, Google Colab, VS Code, Advance Excel

**Machine Learning & AI:**  
Supervised & Unsupervised Learning, Regression, Classification, Clustering, Time Series Forecasting, Recommendation Systems, Feature Engineering, Model Optimization, Deep Learning, Artificial Intelligence, Generative AI, Agentic AI, LLMs (Large Language Models), Transfer Learning, Prompt Engineering

**NLP (Natural Language Processing):**  
Text Classification, Sentiment Analysis, Topic Modeling, Named Entity Recognition, Embeddings (Word2Vec, GloVe, BERT), Tokenization, LLM Fine-Tuning

**Frameworks & Libraries:**  
scikit-learn, XGBoost, LightGBM, TensorFlow, Keras, PyTorch, Hugging Face Transformers, OpenCV, Pandas, NumPy, Matplotlib, Seaborn

**Data Engineering & MLOps:**  
ETL, CI/CD Pipelines, Docker, REST APIs, Real-time Inference, Model Deployment & Monitoring

**Cloud & Platforms:**  
AWS (S3, EC2, SageMaker), Databricks, Kafka, Hadoop, Microsoft Azure, Google Cloud Platform (Vertex AI, BigQuery),

**Tools & Visualization:**  
Power BI, Tableau, Excel, Postman

**Version Control & Project Management:**  
Git, GitHub, Version Control Systems, Agile Methodologies, Project Management

**Statistics & Experimentation:**  
Probability, A/B Testing, Statistical Inference, Hypothesis Testing, Experiment Design, Data Analysis

**Professional Experience**

**Data Scientist | TECEON Software Private Limited | Full-time**  *Dec 2024 – Present (6 months)*

* Designed and implemented **scalable ETL pipelines** using Python and SQL to automate data ingestion, transformation, and validation across multiple sources, leading to a **60% improvement in reporting efficiency**.
* Developed and deployed **machine learning models** for key business use cases including **customer churn prediction**, **sales forecasting**, and **inventory optimization**, leveraging tools such as **scikit-learn**, **XGBoost**, and **Pandas**.
* Integrated ML model outputs into operational systems and **BI dashboards** to enable **real-time data-driven decision-making** across business functions.
* Built and deployed **RESTful APIs and microservices** using Django, enabling seamless interaction between ML models, data dashboards, and external applications.
* Established **CI/CD pipelines** for automated testing, deployment, and version control of ML workflows, ensuring reliability, reproducibility, and maintainability.
* Created interactive dashboards using **Power BI** and **Tableau**, visualizing real-time KPIs and trends by integrating backend data sources for business teams.
* Collaborated in **Agile sprints** with cross-functional stakeholders including product managers, engineers, and business analysts to define KPIs, improve data quality, and align solutions with organizational goals.
* Conducted **exploratory data analysis (EDA)** to uncover insights, guide feature engineering, and support model development.

**Key Skills:**  
Python · Django · SQL · ETL · Exploratory Data Analysis (EDA) · Machine Learning · Deep Learning · XGBoost · CI/CD · REST APIs · Power BI · Tableau · Agile

**Consultant Operations | PolicyBazaar** | **Full-time** *Jan 2021 – Dec 2024 (4 years)*

* Developed and deployed machine learning models for **lead scoring**, **customer segmentation**, **churn prediction**, **pricing optimization**, and **fraud detection**, directly driving key business outcomes.
* Analyzed large-scale structured and unstructured data from **CRM systems**, **web/app logs**, **user behavior analytics**, and **sales funnels** to generate actionable insights and support strategic initiatives.
* Built and enhanced **recommendation systems** using user demographics, behavioral patterns, and intent signals to **personalize insurance offerings** and increase customer engagement.
* Applied advanced **Natural Language Processing (NLP)** techniques on **call center transcripts**, **chat logs**, and **emails** for **sentiment analysis**, **intent detection**, and **automated complaint classification**.
* Designed, executed, and analyzed **A/B experiments** to optimize product features and marketing strategies, contributing to measurable improvements in conversion and retention metrics.
* Partnered with data engineering teams to develop robust **data pipelines**, ensuring consistent data flow and quality for analytics and ML consumption.
* Delivered **real-time business insights** through interactive dashboards in **Power BI** and **Tableau**, enabling data-driven decisions across **marketing**, **sales**, and **operations**.
* Translated complex business requirements into scalable analytical solutions, acting as a key liaison between **data teams**, **product managers**, and **executive leadership**.
* Regularly presented **data-driven recommendations** to senior stakeholders, influencing key decisions in product development, marketing campaigns, and operational processes.

**Key Achievements:**

* Increased lead-to-sale conversion rate by **18%** using a real-time **behavior-based lead prioritization** model.
* Reduced manual review efforts by **30%** through an **ML-based anomaly detection system** for dynamic insurance pricing.
* Boosted **customer retention** via a churn prediction model with **82% precision**, enabling proactive engagement and renewal campaigns.
* Enhanced customer service efficiency by **25%** using **NLP-based complaint classification** and **sentiment analysis**.
* Enabled faster business decision-making by deploying dynamic **real-time dashboards** with predictive KPIs across departments.

**Key Skills & Tools:**  
Python · SQL · scikit-learn · XGBoost · NLTK · spaCy · Power BI · Tableau · Matplotlib · Seaborn · Google Cloud Platform · AWS · Google Colab · Jupyter Notebook · Django · Git · GitHub · Docker · Machine Learning · Deep Learning · NLP · A/B Testing · Recommendation Systems · Data Pipelines · Business Analytics

**Data Analyst | BOTEX AI Private Limited | Full-time**  *July 2018 – Nov 2020 (2.5 years)*

* Analyzed large and complex insurance datasets to uncover business trends, risks, and growth opportunities.
* Built statistical models to forecast customer behavior, policy churn, and claim probabilities.
* Created interactive dashboards and reports to present key metrics and actionable insights to stakeholders.
* Wrote optimized SQL queries to extract, manipulate, and transform data from relational databases.
* Used Python for automation, data cleaning, and generating visualizations for business presentations.
* Developed and maintained robust data pipelines ensuring data quality, consistency, and scalability.
* Segmented customer data to support targeted marketing and personalized user experiences.
* Conducted A/B tests and campaign analysis to evaluate performance and guide business strategies.
* Partnered with product, marketing, and finance teams to understand requirements and deliver data-driven solutions.

**Tools & Technologies:**SQL | Python | Pandas | NumPy | Matplotlib | Seaborn | Power BI | Excel | Jupyter | Tableau

**Impact:**• Improved business decision-making by delivering accurate, timely, and insight-rich reports.  
• Enhanced operational efficiency through automation and predictive analytics in key business areas.

**Education**

**Bachelor of Engineering (Mechanical Engineering)**  
*Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal*  
2014 – 2018 | CGPA: 7.4/10

**Relevant Coursework:**  
Engineering Mathematics · Applied Statistics · Computational Methods · Data Structures · Operations Research

**Academic Projects:**

**Predictive Maintenance Using Machine Learning**  
Built ML models using Python to predict equipment failure from sensor data (temperature, vibration, pressure). Applied Random Forest and regression algorithms to forecast RUL and prevent breakdowns. Improved maintenance efficiency by 25%.

**Inventory Optimization Model**  
Developed a Python and Excel-based model to forecast demand using ARIMA and optimize inventory levels. Reduced holding costs by ~18% and improved product availability by 22% through real-time tracking and replenishment planning.

**ATV Design & Fabrication – SAE BAJA India**  
Led suspension and steering system design for an all-terrain vehicle. Performed FEA on chassis components using ANSYS and managed end-to-end fabrication. Successfully qualified and competed at the national SAE BAJA event.

**Higher Secondary (Class XII)  
2014**  
*Government School of Excellence, Morena, Madhya Pradesh*  
Specialization: Physics, Chemistry, Mathematics (PCM)

**High School (Class X)  
2011**  
*City Montessori Higher Secondary School*  
Secured First Division with Distinction

**Projects**

**1. AI-Powered Investment Analysis Platform**

**Company:** Policybazaar (Offsite)  
**Role:** Consultant  
**Duration:** Jan 2024 – Sep 2024

**Objective:**  
To assist investment teams in evaluating potential deals by building an AI-driven tool that analyzes financial documents and provides actionable insights.

**Tools & Tech Used:**  
Lang Graph, Python, Lang Chain, LLMs, Pandas, Matplotlib, Seaborn, Streamlit, SQL, GCP

**Work Done:**

* Developed an AI pipeline using Lang Graph and LLMs to extract financial KPIs such as EBITDA, CAGR, and revenue forecasts from diverse financial documents.
* Built an intuitive front-end using Streamlit to display results in interactive visual formats (charts, tables, summaries).
* Integrated secure, multi-deal management system with authentication and document versioning.
* Designed a Q&A interface powered by LLMs, allowing real-time natural language querying of deal data.
* Linked the tool with a GCP-hosted SQL database to store structured deal data and enable collaborative access across teams.

**Impact:**

* Reduced document review time by ~60%, enabling analysts to focus on decision-making rather than data extraction.
* Standardized financial analysis across investment sectors.
* Improved collaboration among finance, legal, and strategy teams via a centralized and intelligent platform.

**2. NLP-Based Complaint Classification & Sentiment Engine**

**Company:** Policybazaar (Offsite)  
**Role:** Consultant  
**Duration:** Feb 2023 – Nov 2023

**Objective:**  
To automate the classification of customer complaints and analyze sentiment trends to improve support quality and turnaround time.

**Tools & Tech Used:**  
Python, NLTK, spaCy, BERT, FastAPI, MongoDB, Docker, Power BI

**Work Done:**

* Processed over 500K+ customer complaints to train a BERT-based classifier for tagging issues into predefined categories.
* Built a custom sentiment analysis engine using NLTK and spaCy to identify tone and mood in support tickets.
* Packaged both models into FastAPI services and deployed them using Docker containers.
* Connected APIs to MongoDB for structured storage and integrated outputs into Power BI dashboards.

**Impact:**

* Reduced manual complaint tagging effort by 60%.
* Improved first-response accuracy and lowered average resolution time by 25%.
* Enabled strategic analysis by identifying recurring pain points through sentiment trends across regions and product lines.

**3. Churn Prediction for Insurance Renewals**

**Company:** Policybazaar (Offsite)  
**Role:** Consultant  
**Duration:** Jan 2022 – Oct 2022

**Objective:**  
To proactively identify customers at risk of not renewing their insurance policy and enable retention campaigns.

**Tools & Tech Used:**  
Python, Pandas, Scikit-learn, Logistic Regression, Random Forest, SQL, Tableau, AWS SageMaker

**Work Done:**

* Merged policy, claims, payment, and interaction data to build a comprehensive customer profile dataset.
* Trained Logistic Regression and Random Forest models to predict churn likelihood based on timing, NPS, and claims behavior.
* **Used SHAP for model explainability to ensure transparency and stakeholder trust.**
* Integrated model predictions with marketing automation tools for targeted SMS/email outreach.

**Impact:**

* Achieved 82% precision in identifying potential churners.
* Reduced churn by 12% in the first quarter post-implementation.
* Added ₹20.1 Cr in annual policy renewal revenue through proactive retention measures.

**4. Intelligent Lead Scoring System Using Machine Learning**

**Company:** Policybazaar (Offsite)  
**Role:** Consultant  
**Duration:** May 2021 – Nov 2021

**Objective:**  
To improve sales funnel efficiency by predicting the likelihood of a lead converting into a customer and enabling targeted follow-ups.

**Tools & Tech Used:**  
Python, Scikit-learn, XGBoost, Random Forest, SQL, Pandas, Airflow, Power BI, AWS S3/EC2

**Work Done:**

* Consolidated data from CRM, clickstream, call center logs, and campaigns.
* Engineered behavioral features such as user engagement, ad interaction, and call responsiveness.
* Trained and evaluated XGBoost and Random Forest models, achieving 87% ROC-AUC.
* Scheduled daily model runs via Airflow and integrated lead scores into the sales CRM via APIs.
* Built Power BI dashboards for real-time monitoring of lead quality and team performance.

**Impact:**

* Improved lead conversion rate by 18%.
* Reduced non-productive call volume by 25%, allowing sales to focus on high-quality leads.
* Enhanced revenue pipeline through better targeting and resource optimization.

**5.Logistics Operations Dashboard & Analysis**

**Company:** BOTEX AI (Offsite)  
**Role:** Consultant  
**Duration:** Aug 2020 – Nov 2020

**Objective:**  
To provide real-time visibility into supply chain performance and delivery KPIs using a centralized dashboard.

**Tools & Tech Used:**  
SQL, Excel VBA, Tableau, Python, Data Cleaning, Logistics KPIs

**Work Done:**

* Gathered data from internal warehouse systems and external logistics vendors.
* Cleaned and transformed data using SQL queries and Excel VBA scripts for automation.
* Built interactive dashboards in Tableau to track delivery metrics including SLA adherence, delays, and failure reasons.
* Implemented threshold-based alerts to notify teams of performance breaches in real time.

**Impact:**

* Increased delivery success rate by 10% by enabling proactive interventions.
* Saved 15+ hours/week by automating operational reporting.
* Provided leadership with live, actionable insights for daily logistics decision-making.

**Certifications**

* **Machine Learning with Scikit-Learn** – *LinkedIn*  
  *Issued: Apr 2025*  
  *Skills:* Scikit-learn, Machine Learning
* **Python (Programming Language)** – *HackerRank*  
  *Issued: Apr 2025 | Credential ID: E3D46BCB7209*  
  *Skills:* Python, Control Flow, Strings, Collections, OOP
* **SQL** – *HackerRank*  
  *Issued: Apr 2025 | Credential ID: e1e9bbd943e9*  
  *Skills:* SQL
* **Data Science Job Simulation** – *British Airways (Forage)*  
  *Issued: Feb 2025*  
  *Skills:* Data Analysis, Real-world Data Science Use Cases
* **Introduction to Data Science** – *Commonwealth Bank (Forage)*  
  *Issued: Feb 2025 | Credential ID: EKF4FqHqEyf8DYJcF*  
  *Skills:* Microsoft Excel, Data Engineering Pipelines
* **Power BI Job Simulation** – *PwC Switzerland (Forage)*  
  *Issued: Feb 2025 | Credential ID: MCxnavFe2psQu3D8M*  
  *Skills:* Power BI, Dashboarding, Business Intelligence
* **MLOps** – *360DigiTMG*  
  *Issued: Dec 2024 | Credential ID: 360DTMGIN32901*  
  *Skills:* MLOps, Model Deployment, Monitoring
* **Data Analytics and Visualization** – *Accenture North America (Forage)*  
  *Issued: Oct 2024 | Credential ID: 9AMjnGfAoi76NYmHX*  
  *Skills:* Data Analytics, Visualization, Business Reporting
* **Canva for Work** – *Canva*  
  *Issued: Apr 2024 | Credential ID: 11def3*  
  *Skills:* Design, Presentation, Data Storytelling