Shivateja Koppuravuri

KS | (913) 489-7966 | [shivatejakoppuravuri@gmail.com](mailto:shivatejakoppuravuri@gmail.com)

**SUMMARY**

Data Scientist with 3+ years of experience bridging raw data and machine learning outputs to drive business insights. Adept at utilizing Python, SQL, and BI tools like Power BI and Tableau to create actionable dashboards and monitor model performance. Proven expertise in healthcare analytics and NLP pipelines, streamlining decision-making and reducing operational costs.

**EDUCATION**

**University of Central Missouri, MO, USA**

*Master's, Computer Science*

**PROFESSIONAL EXPERIENCE**

**Blue Cross Blue Shield Aug 2024**

*Data Scientist/Developer KS*

* Analyzed EDI 837 claims using XGBoost and Isolation Forest models to interpret classification outputs, eliminate over 17,000 false fraud alerts annually, and save $850,000 in payout errors.
* Designed NLP workflows (spaCy, BERT) to classify over 500,000 EOB documents monthly, reducing manual classification time and ensuring actionable insights.
* Optimized Oracle SQL pipelines with PySpark on Azure Databricks to reduce eligibility-check processing from 6 hours to under 45 minutes, leveraging skills in SQL and Python.
* Developed Prophet-based forecasts for weekly claims volumes which informed staffing adjustments and significantly reduced monthly backlogs.
* Deployed TensorFlow risk-scoring models in prior authorization workflows, automating decisions for 12,000+ claims weekly and enhancing model performance monitoring.
* Built and maintained Power BI dashboards tracking key metrics—fraud-alert rates, processing times, and claim volumes—providing clear visualizations aligned with business KPIs.
* Authored complex SQL queries and ETL scripts to consolidate claims, member, and provider data into a centralized analytics mart, supporting robust exploratory data analysis.

**McKesson Jan 2024 - May 2024**

*AI Intern KS*

* Developed a PyTorch-based biomedical embedding pipeline to identify over 12,000 high-risk prescriptions monthly, supporting model output analysis and validation.
* Built LSTM demand-forecast models on Kafka streams that fed daily inventory dashboards, effectively aligning model predictions with business KPIs.
* Created XGBoost readmission-risk classifiers (AUC 0.92) to generate weekly patient-risk reports, enhancing collaborative efforts for targeted follow-up interventions.
* Automated dosage extraction from unstructured clinical notes with BERT-NER, efficiently populating EMR tables across 30+ hospital systems while supporting data normalization.

**KPMG Sep 2021 - Nov 2022**

*Data Scientist India*

* Collaborated on fraud-detection analytics over 4 TB of transaction logs using TensorFlow transformers, analyzing model outputs to expose $3.8 million in potential losses.
* Engineered client risk-scoring pipelines with XGBoost and spaCy NLP on SEC filings to unlock over 7,000 audit preparation hours across 200 clients, integrating structured data analysis.
* Developed LDA-driven topic extraction for 10,000+ audit documents, powering Plotly dashboards that visualized 50 emerging risk areas in alignment with strategic KPIs.
* Optimized PySpark clustering workflows to segment clients by compliance behavior, recapturing 200 analyst hours every month and refining predictive grouping.
* Crafted Power BI dashboards consolidating monthly audit KPIs—cycle times, exception counts, cost variances—for senior leadership, ensuring clear business intelligence.
* Constructed SQL ETL pipelines ingesting data from five key systems into a central warehouse, facilitating ad hoc analysis and enhanced data monitoring.
* Supported development of an autoencoder-based anomaly detector in TensorFlow to flag 1,500 unusual expense patterns per quarter, aiding in proactive model refinement.
* Assisted in architecting Prophet time-series models to forecast monthly audit workloads with high accuracy, sharpening resource planning and validation experiments.

**AirBnb**

*Data Science Intern*

**Mar 2021 - Aug 2021**

*India*

* Created XGBoost-based pricing optimization models that incorporated review sentiment analysis (VADER), delivering revenue uplift for over 2,000 hosts and aligning AI outputs with business objectives.
* Deployed collaborative filtering systems using Surprise to develop personalized property recommendations for new users, driving improvements in conversion rates.
* Conducted spatial clustering using DBSCAN on host location data to identify high-demand rental areas, supporting business devel- opment with actionable insights.
* Automated host performance reports by fine-tuning BERT for sentiment extraction across 500,000+ reviews, furnishing scalable and clear visualizations for operations teams.

**TECHNICAL SKILLS**

* **Programming**: Python (NumPy, Pandas, Matplotlib, Seaborn, SciPy, Scikit Learn, TensorFlow, PyTorch), SQL, R (ggplot2, dplyr)
* **Machine Learning**: Regression Analysis, Bayesian Methods, Decision Trees, Random Forests, Support Vector Machines (SVM), Neural Networks, Sentiment Analysis, K-Means Clustering, KNN, NLP
* **Data Visualization**: Plotly, Tableau, Power BI, Advanced Excel (Pivot Tables, VLOOKUP, Macros)
* **Big Data Technologies**: AWS (S3, EMR, Redshift, Lambda, EC2), Azure, Google Cloud Platform (GCP)
* **Databases**: MySQL, PostgreSQL, Oracle DB, MongoDB, Microsoft SQL Server
* **Statistical & Predictive Analysis**: Time Series Forecasting, Predictive Modeling, Hypothesis Testing, Classification, Clustering, Principal Component Analysis (PCA), Deep Learning, Model Evaluation
* **DevOps & Tools**: Docker, Git, Github, Kubernetes, CI/CD, Jenkins
* **Methodologies**: SDLC, Agile (Scrum, Kanban), Waterfall
* **Domain Expertise**: AI Governance Principles, Healthcare Background

**CERTIFICATION**

* AWS Certified Solutions Architect - Associate
* HackerRank SQL Certification
* Python for Data Science Certification