Uma Sivakumar

Austin,TX | umas0697@gmail.com | 979-344-3623 | LinkedIn | GitHub | Portfolio

SUMMARY

Data Scientist specializing in machine learning, predictive analytics, and scalable data solutions to drive business impact. Experienced in building and deploying ML models, optimizing algorithms, and developing data pipelines. Adept at translating complex data into actionable insights, improving efficiency, and enabling data-driven decision-making.

EDUCATION

Texas A&M University — Texas, United States
Master of Science in Data Science (GPA: 3.8/4) — **Dec 2024**

International Institute of Information Technology (IIITB) — Bangalore, India
Certification in Advanced Data Science (GPA: 3.71/4) — **May 2023**

SRM Institute of Science and Technology — Chennai, India
Bachelor of Technology in Information Technology (GPA: 8.72/10) — **May 2019**

SKILLS

Programming Languages: Python, SQL (Databases: Oracle, PostgreSQL, MySQL), R, C++, Java
Machine Learning & AI: Scikit-learn, TensorFlow, PyTorch, Hugging Face, NLTK, OpenCV, Pandas, NumPy, Neural Networks
Cloud, Big Data & Visualization: AWS, Apache Kafka, Spark, HBase, Tableau, Flask, Docker, Git, GitHub, Microsoft Excel, Jira
General Skills: Communication Skills, Root Cause Analysis, Statistical Modelling, Forecasting, Agile Methodologies, Power BI, Data Visualization, Deep learning, Large Language Models, NLP, Data Modelling, Postman

EXPERIENCE

Research Assistant, Texas A&M University, College Station, Texas — **Feb 2024 – May 2024**
• Formulated risk assessment framework using Bayesian network models for US healthcare supply chain utilized by U.S. DHS. Analyzed **300+** variables to inform data-driven decisions to stakeholders, US Department of Homeland Security.
• Applied statistical methodologies to extract actionable insights to curate a **3-node model**, collaborating with cross-functional teams to inform critical decisions.

Software Engineer, Siemens Technology, Bangalore, India — **Jun 2022 – Jul 2023**
• Enhanced system performance by **15%** through development of 6 key products using C++, MFC, and SQL.
• Worked with cross-functional teams to resolve critical performance issues in WinCC Graphics, improving stability by **10%**.
• Conducted reviews of **50+** code submissions, effectively identifying errors and ensuring improved code quality and consistency.

Software Engineer II, Oracle Health, Bangalore, India — **Jul 2019 – Jun 2022**
• Revamped the EHR Millennium Scheduling application boosting usage by **25%** in cancer hospitals.
• Migrated medical insurance tools from VB to C++, boosting performance by **80%**, benefiting clients in **35** countries.
• Designed and documented technical architectures, improving stakeholder clarity and accelerating project delivery by **15%**.
• Mentored freshers and experienced employees on the Scheduling Millennium application driving the team's productivity by **50%**.

Data Intern, Oracle Health, Bangalore, India — **Jul 2019 – Jun 2022**
• Spearheaded the migration of healthcare batch systems to near-real-time scalable ETL pipelines, ensuring **100%** data integrity.
• Deployed a robust transition pipeline leveraging Apache Kafka, Spark, HBase, and Oracle with Java ensuring seamless scalability.

PROJECTS

P.E.E.R: Designed and deployed a personalized recommendation system using BERT embeddings and PostgreSQL GIN indexing on AWS. Handled 160K+ records with dynamic click-based re-ranking.

Translation Uncertainty [Link] **:** Built **3 novel** uncertainty metrics (geometric mean, arithmetic mean, arithmetic mean of scaled-kurtosis) for evaluating LLM translation uncertainty using T5 models.
• Designed a **Flask-based web app** with gradient color schemes to visualize token-level confidence, aiding transparency in generative AI outputs.

Bike Sharing Demand: Developed a predictive model using advanced regularization techniques, achieving **100%** accuracy in forecasting bike-sharing demand and providing actionable insights into user behavior.
• Conducted data profiling and ETL processes, validated using **ANOVA** and **T-tests**, and delivered strategic business decisions.

Maximized Customer Retention in the Telecom Industry: Built Random Forest and XGBoost models to predict customer churn with **91%** accuracy, delivering actionable insights to improve retention strategies.
• Leveraged predictive modeling on large-scale datasets to drive strategic decisions, enhancing business outcome.

AWARDS

• **Night on The Town Award**, Oracle Health: Outstanding 24-hour project turnaround performance received at the Global Townhall Leadership event (Mar 2021).
• **Quarterly Quality Award**, Oracle Health: Recognized for process efficiency and delivery quality (Q1 2021).