

Sandhya Kilari

📞 +1 (517) 580-2675 ✉ kilarisa@msu.edu 💻 in/sandhya-kilari/ 🌐 github.com/SandhyaKilari 🏠 3839 Hunsaker Dr, East Lansing, MI

PROFESSIONAL SUMMARY

Data-driven professional with a robust foundation in machine learning, statistical analysis, and data visualization, skilled at transforming complex datasets into strategic insights. Proficient in building predictive models, implementing scalable ETL pipelines, and optimizing data workflows to enhance business outcomes. Expertise spans advanced tools and platforms, including Python, SQL, Power BI, and cloud technologies (AWS, GCP, Azure), applied across diverse industries and challenges. Proven experience in time-series forecasting, portfolio optimization, A/B testing, and segmentation analysis, with a track record of delivering measurable improvements in decision-making processes. Adept at collaborating with cross-functional teams to design innovative solutions that support organizational efficiency and strategic planning.

TECHNICAL SKILLS

- **Languages:** Python, R, C/C++, ASP.NET MVC, C#, JavaScript, CSS, SQL, HTML, Java, MATLAB, NoSQL
- **Tools and Technologies:** RStudio, Jupyter Notebook, SQL Server, Big Data Tools (PySpark, Hadoop), Power BI, Tableau, Altair, Matplotlib, NumPy, SciPy, Seaborn, Scikit-learn, Pandas, PyTorch, Keras, TensorFlow, Streamlit, Flask, OpenCV, SpaCy, NLTK, AWS, GCP, Microsoft Azure, Visual Studio, Excel (Power Pivot, Power Query)
- **Skills:** Machine Learning, Data Visualization, Data Collection and Manipulation, Predictive Modeling, Natural Language Processing, Deep Learning and Neural Networks, Data Scraping, API Integration, Data Engineering, Functional Data Analysis, Statistical Modeling, Time Series Forecasting, Statistical Analysis, ETL Pipelines, Data Transformation, Artificial Intelligence, Computer Vision (Image Processing, Object Detection, Image Classification, 3D Object Recognition, Scene Analysis, Image Segmentation), Optimization Techniques, Risk Management, Data Analysis, Financial Insights, Model Optimization, Comparative Analysis, A/B Testing, Business Analyst.

EXPERIENCE

Michigan State University

Student Research Assistant III

East Lansing, MI

May 2024 – Dec 2024

- **Health Utilization Prediction for Cancer Patients:** Developed machine learning models to predict **120-day survival and emergency room utilization** for cancer patients by integrating phenotypic data, cancer risk factors, Social Determinants of Health (SDOH), and Patient-Reported Outcomes (PROMs). Applied advanced analytics techniques to identify significant predictors like the Charlson Comorbidity Index (CCI), achieving an **AUC of 0.92**. Streamlined data handling processes by addressing missing data challenges, enhancing the reliability of the predictive models. Delivered actionable insights to support **clinical decision-making** and improve patient care outcomes
- **Dry Bean Quality Prediction:** Utilized **Vis-NIR spectroscopy data** and **protein content metrics** to develop **machine learning models** (SVM, ANN) for predicting dry bean quality traits, including **color, texture, size, and protein concentration**. Implemented **dimensionality reduction** and **A/B testing** to uncover relationships between categorical variables, enhancing model accuracy. Key data analysis of **particle density** and **size distribution percentiles** led to **92% prediction accuracy** across genotypes and reduced manual inspection time by **40%**, improving **quality assessment reliability** and enabling **data-driven decisions** in breeding programs.

QSIDE Institute (Capstone Project)

Internship (Research and Development Fellowship)

Williamstown, MA (Remote)

July 2024 – Nov 2024

- Created and Implemented comprehensive **data lake using AWS S3** to centralize and store grant opportunities for Native American tribes and Nonprofits, streamlining the application process for federal, foundation, and corporate grants
- Utilized **GCP's Gmail API** and **API calls** to automate the retrieval and scraping of grant data from newsletters, databases (e.g., Grants.gov, JustGrants, Instrument1), and IRS PDFs, achieving a **20-30% improvement in accuracy** through custom **AI models using OpenAI** for filtering, validating and classifying grant-opportunities based on eligible criteria. Automated data retrieval and reporting processes using **Tableau**
- Developed a user-friendly interface that **automated updates** and delivered actionable insights, improving **grant discovery accuracy** and **data management efficiency**. Presented data findings to **non-technical stakeholders**, translating insights into recommendations that **influenced business decisions**

Accenture Solutions Pvt. Ltd

Analyst

Bengaluru, India

Nov 2020 – July 2023

- Streamlined data migration processes, password management, upgrades, and testing for three applications in line with business requirements using **C#, SQL Server, .NET, JavaScript**, and **Azure DevOps**, ensuring smooth data transitions, improved security protocols, and maintained overall system stability by optimizing **data pipelines**
- Managed a data-driven project for fund allocation on the **Azure Platform** through **Dynamics 365 CRM**. Validated and processed large datasets, ensuring compliance and data accuracy across multiple program types. Increased processing **efficiency by 20%** through workflow automation, reducing manual input and errors
- Earned **Power Platform Fundamentals** certification, gaining expertise in building data-driven solutions and automating workflows using the **Power Platform**, further enhancing efficiency in managing and improving application processes

EDUCATION

Michigan State University

Master of Science in Data Science | GPA: 3.9/4.0

East Lansing, MI

August 2023 – May 2025

Relevant Coursework: AI Computation Foundations, Data Mining, Statistical Modeling, Big Data Analysis, Computer Vision, Computational Optimization

PROJECTS

- **Heart Disease Risk Assessment:** Created a predictive model to assess **heart disease risk** using the **Kaggle Heart Disease dataset**. Applied Random Forest, K-Nearest Neighbors, and Support Vector Machine algorithms, achieving **95% accuracy** with Random Forest. Designed an interactive **Streamlit web application** for easy user risk assessment, utilizing health attributes like age, gender, cholesterol levels, and blood pressure
- **Predicting Credit Default:** Forecasted customer credit defaults by leveraging historical payment data and demographic information from the Kaggle Default of Credit Card Clients dataset. Conducted **A/B testing** to evaluate model performance across multiple algorithms, including **Random Forest, XGBoost, MLP, and SVM**, with **Random Forest** achieving **93% accuracy**. Optimized features using **Principal Component Analysis (PCA)** and addressed class imbalance with oversampling, improving **credit allocation** and mitigating **default risks**
- **Bike Rental Prediction Analysis:** Constructed predictive models to forecast bike rentals based on **weather** and **temporal data** using multiple linear regression and multi-layer neural networks. Leveraged **Keras** and **TensorFlow**, achieving a **RMSE of 335.36**, outperforming linear regression (**RMSE: 442.23**). Experimented with **PyTorch** and identified key predictors like temperature and time of day
- **Sentiment Analysis:** Engineered a **sentiment analysis system** to classify customer emotions from **Google reviews** using Naive Bayes, Logistic Regression, and Stochastic Gradient Descent (SGD) models. Utilized **TF-IDF, Bag of Words, and SpaCy** for vectorization, with **NLTK** for text preprocessing, enhancing data quality and model accuracy. Deployed the system into a **Flask-based web application** for real-time sentiment analysis, achieving **94.6% accuracy** with the SGD model
- **Graph-Based Node Classification:** Developed a **Graph Convolutional Network (GCN)** achieving **89% accuracy** on a dataset of 2,480 nodes and 10,100 edges. Optimized model performance using **Bayesian optimization**, with precision reaching **1.00** for some classes and addressing recall challenges
- **Functional Graphical Models for Time-Series Data:** Implemented **nonparametric high-dimensional** functional graphical models using Additive Functional Graphical Models (AFGM) and Functional Gaussian Graphical Models (FGGM). Leveraged techniques like **Functional Principal Component Analysis (FPCA)** and **B-splines** for analyzing time-series data, achieving an **AUC of up to 88%** in predictive performance
- **Portfolio Optimization with Constrained Optimization Techniques:** Developed a portfolio optimization model using Sequential Least Squares Programming (SLSQP), achieving a **27% improvement in Sharpe Ratio** (1.362 vs. 1.070) over traditional methods. Conducted rolling-window backtesting for dynamic market adaptation and applied advanced risk metrics like **CVaR** and **MDD**, reducing drawdowns by up to 69%. Leveraged Python, Scipy, and Pandas to analyze a **diversified asset portfolio**, enhancing **risk-return balance** and **predictive performance**
- **Heart CT Segmentation with Advanced Deep Learning Models:** Developed and evaluated **segmentation models** for heart CT images, achieving a Dice Score of up to 92% with Attention U-Net. Compared state-of-the-art architectures, including PSPNet, U-Net, SegNet, and a custom CNN, to enhance segmentation accuracy and **computational efficiency**. Leveraged advanced preprocessing techniques (normalization, noise reduction, augmentation) and implemented innovative features like attention gates and pyramid pooling for superior segmentation performance, demonstrating the potential of deep learning in **clinical workflows**.

LICENSES & CERTIFICATIONS

- **4th Annual Henry Ford + MSU Cancer Research Symposium (2024):** Presented research work on cancer prediction
- **QSIDE Institute Datathon4Justice Leader (2024):** Recognized for leading impactful data-driven projects for social justice
- **GCP for Clinical Trials Investigators and Staff:** Certified in Good Clinical Practice standards for clinical trial management
- **Microsoft Certified Power Platform Fundamentals:** Proficient in building data-driven solutions using Microsoft Power Platform