Karthik Kodakandla

kkodakandla@horizon.csueastbay.edu | +15104325384 | LinkedIn | GitHub

SUMMARY

Data science enthusiast with a master's in business analytics (data science major) having around 7 years of experience in analyzing complex and large datasets, creating data models with BI tools (ETL), providing business insights through reporting tools. Proficient in predictive modeling, data processing, feature engineering and data mining algorithms using machine learning, deep learning, SQL, and python. Seeking opportunities for data scientist roles.

EDUCATION

MS in Business Analytics (Data Science) from California State University - Eastbay (GPA: 3.96/4) **Jan 2020 - Dec 2021**

Course work: Machine Learning, Deep Learning, NLP, Data Mining, Statistics, Big Data, Data Warehousing, and Business Intelligence, Optimization.

Bachelor's in Information Technology from Jawaharlal Nehru Technological University (GPA: 7.5/10) **Sep 2007 - May 2011**

TECHNICAL SKILLS

Machine Learning and Deep Learning: Logistic Regression, Linear Regression, Support vector machine, KNN, Naive Bayes, Decision trees, Xgboost, Random Forest, PCA, K -Means, Association rule mining, Collaborative filtering, Neural Networks, Encoder-Decoder, NLP

Programming: Python, R, SAP ABAP, SQL.

Frameworks and Libraries: Pandas, NumPy, Matplotlib, TensorFlow, Kera's, PyTorch, NLTK, Scikit Learn, OpenCV, Streamlit, Heroku, Flask, Airflow

Visualization Tools: Power BI, Tableau, BEX Query Designer, SAP Business Objects

Database/DW/Big Data: MYSQL, PostgreSQL, SAP BI/BW SAP HANA, PostgreSQL Hadoop, Pig, Hive

Tools and Technologies: SSMS, Jupiter Notebook, GitHub, Docker, Kubernetes, PyCharm, Visual Studio, Jira, AWS, Azure, Solver Table.

PROFESSIONAL EXPERIENCE

Data Science Intern, Ecolab **Jun 2021 - Aug 2021**

- Created aggregated customer dataset by data blending from sources such as CRM SQL server and SAP CRM using Alteryx designer.
- Developed statistical ML models using logistic, CART, random forest, DNN to anticipate account health and evaluated best model with 72% AUC.
- Recommended top products and provided insights using web app(streamlit) and dashboard (PowerBI), increasing customer retention by 3%.

Data Engineer, Royal Dutch Shell **Aug 2017 - Dec 2019**

- Implemented statistical ML models to forecast max discounts, improved model performance and reduced training time by 8% using xgboost.
- Interpreted ML models using SHAP and LIME python libraries to present actionable insights which increased 3-5% annual profits to the business.
- Built operational interactive dashboard and reconciliation reports using power BI and SAP BO. Analyzed profit & loss, balance sheet and AR/AP accounts with business logic and saved approx. 200-man hours of manual efforts per month.
- Designed, developed, and deployed dimensional models (ETL) and data pipelines using SAP BW/HANA in sprints (Certified scrum master).
- Enhanced user experience by performance-tuning HANA SQL views for efficient data retrieval from 40M records accessed by 5000+ users.
- Led a team of new graduates to support UAT and regression testing before deploying to production, reducing hyper care defects by 75%.

Business Intelligence Engineer, Medtronic **Jan 2017 - Jul 2017**

- Assessed performance analysis and optimized solutions for data models and BI reports which improved overall efficiency by 15%.
- Envisioned and enhanced new features in existing ETL models and reports and deployed them to production with 0% error rates.

Business Intelligence Analyst, IBM **Oct 2012 - Jan 2017**

- Analyzed, resolved and found RCA of SAP BW prod issues related to financial closures as an application support team member with 99% SLA.
- Automated Jobs for Data Pre-Process, Data Conversions, Data cleansing, Data Manipulation and saved 6-man hours per day.
- Built data models by gathering requirements from end-users, assessed them to publish High-level Design Documents.
- Provided business insights to the customers by analyzing raw data which helped them to make data-driven decisions.

ACADEMIC PROJECTS

IMDB Score prediction: [***Data Mining| Python| R| Beautiful Soup***]

- Extracted, preprocessed and selected important features of around 5000+ movie data from IMDB using python and beautifulsoup.
- Generated regression model with MAPE to foretell rating and classification model with 81% accuracy to forecast categories of IMDB rating.
- Determined clustering analysis through K-mean algorithm to see major factors contributing to the movie rating.

E-commerce Women Clothing Product Recommendation: ***[Natural Language Processing| NLTK | TF-IDF | Sentiment Analysis]***

- Pre-processed user reviews data and explored text data by visualizing it in the word cloud.
- Devised pipeline model using TFIDF Vectorizer, word2vec and multi-class classifier to predict product recommendations with 88% accuracy.
- Performed sentiment analysis on customer reviews using Vader sentiment analyzer.

Cassava leaf disease classification (Kaggle): [*PyTorch | EfficientNet-B3 | Albumentation | Regularization Techniques*]

- Classified each cassava image into four disease categories or a fifth category indicating a healthy leaf.
- Boosted imbalance classes, tuned hyperparameters and applied regularization methods to improve accuracy by 15% from baseline performance
- Used PyTorch framework and Efficient Net B3 class weights to fine-tune the model and achieved 88.7% accuracy (Top 30% in leader board).

Automatic Numberplate detection and recognition: [*Python |Open CV| Tesseract OCR*]

- Detected Licence plate from an input image by estimating bounding boxes using transfer learning (VGG16) with a 0.8 IOU score.
- Cropped Licence plate from the image using bounding boxes which are then segmented on characters using open CV contours.
- Segmented characters are passed to anticipate the numbers and characters using MobileNetV2 and Tesseract (OCR) with 89% accuracy.