Karthik Kodakandla

kkodakandla@horizon.csueastbay.edu | 5104325384 |San Diego, CA |LinkedIn |GitHub

SUMMARY

Data science enthusiast with a master's in business analytics (data science major) having 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, 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, Oracle, PostgreSQL Hadoop, Pig, Hive
Tools and Technologies:	SSMS, Jupiter Notebook, Spyder, GitHub, Docker, Kubernetes, PyCharm, Visual Studio, Jira Amazon EMR, S3, Redshift, Microsoft Azure, Service Now, HP ALM, Solver, Solver Table, Agile Methodologies

PROFESSIONAL EXPERIENCE:

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

- Created aggregated customer dataset from multiple sources such as CRM SQL master data, SAP CRM transaction data.
- Developed statistical machine learning models using logistic regression, decision tree, random forest, and deep neural networks to predict customer account health and evaluated the best model. Achieved 72% AUC for the best model evaluated.
- Recommended top 5 products to the customers using collaborative filtering based on customer infographics.
- Provided insights from the ML models to the stakeholders using web app(streamlit) and dashboard (PowerBI).

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

- Implemented ML models to predict maximum discounts that can be offered to aviation customers. Improved accuracy by 5% and reduced training time by 8% using xgboost algorithm which increased 3-5% annual profits to the business.
- Interpreted ML models using SHAP and LIME python libraries to present actionable insights to the business.
- Created operational interactive dashboard and reconciliation reports using power BI and SAP BO to analyze profit & loss, balance sheet and AR/AP accounts with business logic. This has saved approx. 200-man hours of manual efforts per month.
- Designed, developed, and deployed dimensional models (ETL) and data pipelines using SAP BW/HANA with an agile framework. (Certified scrum master)
- Enhanced user experience by performance-tuning HANA SQL views for efficient data retrieval from 40M records accessed by 5000+ users worldwide.
- Managed and supervised a team of newly hired graduates to support UAT and regression testing before deploying solutions to production.

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

- Evaluated performance analysis and optimized solutions for data models and BI reports which improved overall efficiency by 15%.
- Developed new features and enhancements to existing models and reports and deployed to production

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

- Analyzed and resolved high-security issues raised during the financial month-end and year-end book closures as a technical support team member.
- 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, analyzed them to publish High-level Design Documents and developed solutions.
- 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***]

- Created regression models (linear regression, KNN) to predict movie rating and classification models (logistic regression, neural networks) to predict categories of IMDB rating (flop, average, hit categorized by binning IMDB score)
- Developed K-mean clustering algorithm to analyze 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 analyzed it by visualizing in the word cloud.
- Created a pipeline model using TFIDF Vectorizer, word2vec and multi-class classifier to predict product recommendations.
- Performed sentiment analysis on customer reviews using Vader sentiment analyzer.

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

- To classify each cassava image into four disease categories or a fifth category indicating a healthy leaf.
- Used class weights to boost imbalanced classes and applied regularization techniques like L1 and L2 to improve 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 predicting bounding boxes using transfer learning (VGG16).
- Cropped Licence plate from the image using bounding boxes which are then segmented on characters using open CV contours
- Segmented characters are passed to predict the numbers and characters using transfer learning (MobileNetV2) and tesseract (OCR)

Awards

- Hackathon Winner 2021 @Ecolab sponsored by Microsoft for developing an analytical prototype for App rationalization using Microsoft products.: Jul 2021
- Shell EPF VP Award for contribution towards a successful development and implementation of a project into production: Jan 2019
- IBM Eminem & Excellence Orion Award for excellence in customer support: Aug 2015