

Shashi Kiran C



Location: Bengaluru, Karnataka, India, 560064

Contact: +91 889-233-9474

Email: c.shashikiran421@gmail.com

Portfolio: <https://kiransview.disha.page>

GitHub: <https://github.com/kiransview>

LinkedIn: <https://www.linkedin.com/in/shashikiranc/>

**Data Scientist | Data Analyst | Machine Learning | Deep Learning | Generative AI
Procurement & SupplyChain Domain | Natural Language Processing | Team Management
Python Programming | Strategic Planning | Time Management | Client Relations
Problem Solving | Forecasting | SQL | AWS | OpenAI | Google Cloud | Major Domo**

Data scientist with 7+ years of experience leveraging statistical modelling, data visualization, data processing, data mining, machine learning algorithms to solve challenging business problems and natural language processing. Total 6+ years of extensive experience in various platforms like data science, Machine Learning and Data Analysis. Successful at the promotion of company advancements and boosting companies' performance. Consistent at reaching revenue targets through leveraging of BI data & Generative AI data modelling and production of well-executed recommendations.

CORE COMPETENCIES

- Experience in Data Analysis, visualization, statistics, programming, and Machine Learning techniques.
- Extensive Model building experience with Machine Learning algorithms for Product.
- Experience in creating mature Data science pipelines encompassing Data standardization, Feature extraction, model validation and optimization.
- Exploring and visualizing data to drive insights.
- Advance knowledge of ML algorithms such as Ensemble Models, Ada-boost, Gradient-Boosting, KNN, Grid search, Random Forest, Decision Tree, SVM, Linear Regression, Ridge, Lasso, Cross Validation, KFold.
- DL algorithms such as ANN, CNN, RNN - LSTM, Transformers, Bert, Seq Model.
- NLP text preprocessing techniques such as Stemming, Lemmatizing, Tokenization, TF-IDF, Bag of Words (BOW), Stop words, Re (Regular Expression), Bigrams, Trigrams, n-grams, Word2Vec.
- Experience in SQL Database with DDL, DML Querying commands.
- Good knowledge and implementation experience in Ensemble algorithms.
- Hyper-parameter tuning to improve the model.
- Participating in constant learning through training and skills development.
- Implementation of Generative AI utilizing the API from the Open AI and generate the Question-and-Answer bot for the document which will be parsed.
- Forecasting Model Build & Monitor for the Demand during various seasonality and tune based on the requirement.
- Systems Utilization and Technology Skills. Proficient use of some Visualization tools such as Tableau & Power BI, with some

PROFESSIONAL EXPERIENCE

SENIOR DATA SCIENTIST

Jul'22 – Present

Aligned Automation Services Private Ltd.

Project: Forecasting the Demand Supply of the Warehouses:

a) EDA & Data Prep for the Model Input:

Technologies: Databricks, Python & its libraries

- Conducted exploratory data analysis to identify demand and supply patterns for warehouse forecasting.
- Cleaned and preprocessed historical warehouse data, ensuring accuracy and reliability for modeling purposes.
- Applied time series analysis techniques to capture trends, seasonality, and autocorrelation in warehouse data.

b) Forecasting Model Build:

Technologies: Databricks, Python, Prophet, ARIMA & Croston

- Developed forecasting models using Databricks, Python, and Prophet to accurately predict warehouse demand and supply.
- Employed ARIMA and Croston models to capture time series patterns and optimize warehouse forecasting accuracy.
- Leveraged advanced technologies and methodologies to build robust forecasting models for efficient warehouse management.

Project: Generative AI:

a) Generative AI using Open AI & Hugging Face Libraries (API):

Technologies: Python, OpenAI, Hugging Face, LangChain, NLP

- Utilized OpenAI and Hugging Face libraries in Python to develop generative AI models for natural language processing (NLP).
- Integrated LangChain for seamless data preprocessing and advanced language modeling in generative AI applications.
- Demonstrated expertise in leveraging cutting-edge technologies to create innovative and efficient generative AI solutions.

Tools Used:

DOMO (Viz Tool), Python, GCP - Vertex AI, Big Query, Teradata, SQL, Tableau, python Libraries such as NumPy, Pandas, Matplotlib, Seaborn, ScikitLearn

DATA SCIENTIST

Feb'21 – Jul'22

Tata Consultancy Services (TCS)

In the Process of improving the Store Performance in different Localities, based on a predictive model.

Project: Store Remodel Performance Analysis:

a) EDA in Improving the Sales Model:

Technologies: DOMO, Python, Jupyter Lab, GCP - Vertex AI, Big Query, Teradata, SQL

Predicting the customer sentiment with the internal Engineer classification labels.

- SQL Data inputs from the Teradata (Store Transactions) used to build the Sales prediction.
- Data Flow preparation using GCP (Jupyter Lab) for some EDA analysis to output the standard reports for the client management requirement using DOMO, Tableau.
- Based on the EDA the auto Dataflow are mapping the routing analysis & prepare an auto refresh model

b) Sales Transference Model:

Technologies: Python, Jupyter Lab, GCP - Vertex AI, Big Query, Teradata, SQL

Predicting the customer movement due to the current closure of the store & their purchase pattern changes to a different store.

- Teradata SQL flows for getting an output of the Sales transference occurred during the occurrence.
- Monthly & weekly transference numbers reporting for the store managers.

c) Market Basket Confidence Analysis:

Technologies: Python, Jupyter Lab, GCP - Vertex AI, Big Query, Teradata, SQL

Predicting the customer sentiment on cross shopping of different products along with the similar mapped products & to generate an insight to the store managers.

- Finding the confidence intervals for the mapping of products.

Tools Used:

DOMO (Viz Tool), Python, GCP - Vertex AI, Big Query, Teradata, SQL, Tableau, python Libraries such as NumPy, Pandas, Matplotlib, Seaborn, ScikitLearn

SENIOR DATA ANALYST

Mar'18 - Oct'20

HCL Technologies

Leading a team of 6, in the improvement of CDQ (Cost, Delivery & Quality) for a Service Wing of a Client.

Project #1: Reduction in Looping cost in logistics Client Service Sector:

a) Customer Sentiment Analysis for a service returned product:

Technologies: TensorFlow, Keras, LSTM, NLP

Predicting the customer sentiment with the internal Engineer classification labels.

- Text input collection from the product team Service engineers, analysis for the loopholes in the data & to get clarity from the engineers (Failure reason).
- Data Preprocessing with Regex & NLTK for preprocessing, TFIDF, W2V & Gensim for Word Embedding the text output.
- Implemented architecture with customer Engineer Sentiment Analysis for Multiclass Labels, using LSTM & Seq2Seq Model to improve the accuracy of sentiment classifier.
- Based on the review with the client, the model was deployed on their SAP architecture to the decision makers.

b) Report auto analysis from the Customer Field Engineer for the failure.

Technologies: Scikit-Learn, Machine Learning Algos, Data Analysis and Visualization

Predicting the customer field engineer with the internal Engineer classification labels.

- Understanding of numerical, categorical and text data using statistical analysis of which includes Univariate Analysis, correlation, statistical distribution of variables, etc.
- Automate the process of text extraction, and attachments from Outlook using python script.
- Annotated, and labeled the text corpus and created a training dataset.
- Pipeline building and Deployment of classifier in SAP server.

Project #2: Cost Analysis for the Repair Products:

Reduce the delivery lead time from 2 tracks, from Supplier & to customer, by utilizing the data from past history from the team.

- Overseeing the collection of the data for prediction from different sources inside the client org.
- Understand the data obtained from the team and then further conduct comprehensive data cleaning and meticulous EDA.

- Predicting the regression output for weekly, monthly data generated from the supply chain team of the client (Based on service products).
- Usage of Machine learning regression models such as Random Forest, SVM, Linear Regression & boosting algorithms to predict the price of the product. Collecting the data from the buyers/Client Logistic Managers.
- Analyzing the lead time using Tableau & to identify the pain points & to highlight the improvement areas.
- Usage of Machine learning regression models such as Random Forest, SVM, Linear Regression & boosting algorithms to predict the price of the product.

Tools Used:

Tableau, PowerBI & python Libraries such as NumPy, Pandas, Matplotlib, Seaborn, ScikitLearn, NLTK, Re.

MANAGEMENT INTERN

Jan'16 – Feb'18

Bosch India

As a Management Intern had some real opportunity to work on some Supply Chain data processing operations.

- Solve technical and data related problems to meet the needs of the Supply Chain business
- Prepare and maintain documentation for reports, metrics, and processes.
- Collecting and interpreting data through Tableau & PowerBI.
- Provide feedback on the effort, risks, and resources necessary on potential project initiatives.
- Examine the needs and concerns of Cross Departmental Leads to determine opportunities or problem areas and provide feedback in the development of policies, procedures, methods and/or standards

Tools Used:

Tableau, PowerBI

INTERN ENGINEER

Jul'15 – Dec'15

Indus Engineering Pvt. Ltd

Great experience in a small-scale industry and my first work experience. Was responsible for Designing Pad printing machines (modelling and drafting) & maintained some of the storage and managing the ledger of inventories.

PERSONAL PROJECTS

REAL NEWS & FAKE NEWS CLASSIFIER (END to END)

(In Research)

As the world is generating millions of News articles/news daily, almost 60 - 80% of News is Fake. To contribute towards the world this small deployment to classify the Fake or Real News using Python (Scikit Learn, NumPy, Pandas, TFIDF), Machine Learning, Natural Language Processing (NLP), Flask & Heroku CLI with GitHub Deployment. Gives a clear classification of the data whether Real News or Fake News.

Problems Encountered (post deployment):

- Bias of the Data (US Region).
- Preprocessing with additional generative options.
- Sequence information is not up to the mark.
- Better UI.

Solutions planned:

- Data Collection & avoiding the bias of the US region.
- Preprocessing of Bi-grams, Tri-Grams, Word2Vec & Word Embedding.
- Algorithm built for this use case provided a good output using Naive Bayes - 98% accuracy, but the sequence information was missing and planned to move with usage of RNN & RNN-LSTM to get the desired outputs.
- Planning for Template remodification using HTML & CSS tags.

Deployment Link: <https://real-fake-news-classifier-mnb.herokuapp.com/>

CAR PRICE PREDICTION (END to END)

(In Research)

Thousands of cars manufactured daily & with many brandings, buying a car is a trend in the recent generation (especially buying used cars). So, predicting a car's price based on the Structure, Make is challenging & year of manufacturing is quite challenging, so by considering some history of sales data & using Python, Machine Learning, Flask & Heroku Deployment. Gives a price prediction on the different make & manufactured year with different features (In research).

Problems Encountered (post deployment):

- Bias of the Data (US Region Cars).
- Preprocessing with additional Combined features.
- Improvement of Model Performance.
- Better UI.

Solutions planned:

- Data Collection & avoiding the bias of the US region.
- Implementation of PCA & generate mean, median of several generalizable features.
- Algorithm built for this use case provided an average output of 85% using Random Forest, planning to hyperparameter tuning of the model with different arguments for the forest model & to use some boosting techniques to the best. More prediction to be done using ANN based on the tuning performance.
- Planning for Template remodification using HTML & CSS tags.

Deployment Link: <https://real-fake-news-classifier-mnb.herokuapp.com/>

EDUCATION AND TRAINING

Liverpool John Moores University (LJMU) Masters of Science in Machine Learning & Artificial Intelligence	(2023 – 2024) (Pursuing)
International Institute of Information Technology Bangalore (IIITB) Executive PG Program in Machine Learning & AI (EPGMLAI)	(2022 – 2023)
PGDM - NMIMS Post Graduate Diploma in Business Management (PGDBM)	(Nov - 2020)
Visvesvaraya Technological University – M.S. Engineering College Bachelor of Engineering - BE, Mechanical Engineering, 67%	(2012 – 2016)
Dept of Technical Education – M.N. Technical Institute High School Diploma, Mechanical Engineering – 64%	(2009 – 2012)