Koshal Kumar (Data Scientist)

+91 8210004520 koshalsingh0921@gmail.com LINKED IN GITHUB Chennai, IN

SUMMARY

Data Scientist with 3+ years of experience in data preprocessing, EDA, data modelling and Machine Learning. Skilled in Python, data wrangling, data structures, inferential statistics, and various data science packages. Worked on projects across multiple industries, such as automation, data modelling, price optimization, data science package development and backend development. Seeking to leverage my expertise and passion for data to create innovative solutions for challenging problems.

PROFESSIONAL EXPERIENCE

Data Scientist @ *Tiger Analytics [* March '22 – Present] Chennai, IN

Price Optimization & Algorithm Development

- Developed an **algorithm** for **revenue forecasting** using the concept of price elasticity of demand for **FMCG client.**
- Deployed multiple **loss minimization** & **optimization techniques** to run examine models and simulation.
- Applied **pricing models** and conducted **pricing simulations** to evaluate the impact of pricing decisions on revenue and profitability.
- Created a method to **resolve the infeasible solution problem and issue of constrained results**.

Dashboard Data

- Developed data cleaning pipelines for several different retailers' datasets using DataBricks, ensuring consistency and accuracy in data processing for a **FMCG client**.
- Implemented quality control measures at each stage of the data cleaning process to validate data integrity and completeness.
- Standardized data formats across all retailers by mapping to a common column (EAN), facilitating seamless data consolidation.
- Orchestrated the consolidation of cleaned data and its subsequent transfer to SQL Pipeline, streamlining data management and analysis processes.

Associate data Scientist @ *Alunoz Technologies* [Dec '21 - Feb '22]

Online E-KYC System

- Responsible for designing and **developing Facial Recognition System**, using open CV's **one-shot learning technique**.
- Worked with agile methodologies.
- Used **DeepFace architecture**. Got Accuracy of 93%.

EDUCATION

M.Sc. Data Science

Central University of Rajasthan [July '19 July' 21] Jaipur, IN

- **Course Modules:**
  - Advances Statistical Methods | Probability Distribution | Linear Algebra and Matrix Theory
  - Python | Machine Learning | Econometrics and Finance | Hypothesis Testing
  - Data Mining | Deep Learning | Natural Language Processing | Bioinformatics | Value Thinking

- **Secured 7.83/10.00 GPA**

BSc Information Technology (Magadh University [Jul '16 - May '19])

- **Secured 73%**

KEY SKILLS

- **Machine Learning Methodologies • Optimization Techniques • Text Mining • Data Mining & Analytics**
- **Predictive & Statistical Modelling • Pricing Analytics • Data Structures • Data Analysis**
- **Data Visualization • Multiprocessing and Multithreading • Predictive Modelling & Analytics • Pricing optimization • Debugging**

TECHNICAL SKILLS

Languages **:** Python, C, Java
Cloud Computing **:** Azure, DataBricks
Machine Learning **:** Sklearn, Regression, Classification, NLP
Database **:** MySQL
Statistics/ML **:** Linear/Logistic Regression, Ensemble Trees, Gradient Boosted trees
Other Skills **:** SQL, Django, Pyspark, Docker, MLflow

KEY DATA SCIENCE PROJECTS

- **Domain: FMCG | Tech Stack: Python, Pyspark, PyOMO, ML, Regression | Jan '24**
  - Objective: A retail price optimization project was conducted to ***develop a pricing strategy that maximizes profit and sales*** while considering various factors such as competition, customer behavior, and market trends.
  - Building a pricing engine that computes optimal prices at the store and item level granularity, optimizing the given objective (e.g., net sales/net margin) by considering several factors, such as demand, cost, inventory, competition, and promotions.
  - Developed a simulation tool that allows business executives to evaluate the impact of price changes on different KPIs (e.g., net sales, margins, and revenue).
  - Key Achievement: Made the ***model 3x-4x faster*** and able to provide actionable insights on price changes to be made that give accurate predictability on the forecast of revenue/margin.

- **Domain: Social Media & Community Platforms | Tech Stack: Python, NLP | Nov '21**
  - **Objective**: Developed a machine learning model to identify duplicate questions on Quora, improving the efficiency of question-answering systems.
  - **Preprocessed** and cleaned text data using **NLP techniques** such as **tokenization, stemming, and stopword removal**.
  - Engineered features like **word embeddings (TF-IDF, Word2Vec)** and **similarity metrics** (cosine similarity, Jaccard similarity).
  - **Trained and evaluated models** (e.g., **Logistic Regression, Random Forest, XGBoost**) to **classify question pairs** as similar or dissimilar.
  - Achieved an **accuracy of 98%** and reduced redundancy in Quora's question database.
  - Utilized **Python libraries** such as **Pandas, Scikit-learn and NLTK**.

CERTIFICATIONS

- Insurance Domain Certification | TA | Oct’ 2023
- CPG Domain Certification | TA | Jan’ 2023
- Applied Natural Language Processing | NPTEL | Dec' 2020
- Core Java Certification | CDAC | Jun' 2017
- Data Structure Certification | CDAC | Sep' 2017