KEVIN PATEL

kevin18patel@gmail.com • (517) 505 9591 • LinkedIn • Portfolio • USA

Results-oriented Data Scientist with over 4+ years of hands-on experience in crafting innovative Data Science, Data Mining, Analytics, Computer Vision and Predictive modeling solutions for enterprise applications. Seeking full-time roles (May-2024).

WORK EXPERIENCE

Data Scientist Capstone | Siemens, Remote, USA — Jan 2024 – Apr 2024

- Aimed at enhancing design pattern parameter predictions for engineering teams in multi-objective optimization problems
- Working on modeling for predicting Features of Perito-fronts (2 objective functions) by optimizing IGT-loss for ZTD-1 dataset using inverse-DNN, Mixture Gaussian Network, Etc.
- Optimized feature generation for the production team to improve the success rate, which prevented a loss worth $220K

Research Assistant | Michigan State University, East Lansing, MI — May 2023 – Apr 2024

- Working on predicting gene-expression changes using 3D gene structural data, Attention-based GNN and DNA-BERT.
- Optimized the code of fetching various gene parameters from 70 minutes to 4 minutes by utilizing SQL query optimization.
- Developed a web dashboard using matplotlib, seaborn, Plotly for conclusive story and business insights for Hospital data analysis and Built a predictive model for predicting the approximate stay length of new patients

AI/ ML Engineer (Gen-AI) Intern | VERN.ai, Remote, USA — May 2023 – Aug 2023

- Integrated Llama 2 (Hugging Face LLM) to collect data for user queries and improve customer satisfaction
- Finetune Llama-2 using instructions using the instruction dataset (Hugging Face) and monitor user preference using A/B testing to reduce toxicity and harmless response generation. Used SFT-trainer and transformer-trainer along with PEFT (LoRA and QLoRA techniques) and Quantization technique keeping BLUE, ROUGE score, and toxicity score into consideration

Data Scientist | India Meteorological Department (IMD), Bhopal, India — Jul 2021 – Aug 2022

- Translated unlabeled RADAR data into actionable insights using Python libraries (scikit-learn, OpenCV), & rectified cloud height measurement errors for a 25% accuracy improvement, enhancing weather data precision
- Enhanced data analysis through dimensionality reduction, feature extraction & interpolation. Achieved a 20% accuracy boost in Central India's weather forecasting while executing LSTM, GRU, ARIMA.
- Implemented Unsupervised Approach for clustering scattered cloud data by k-means, DB-scan, k-medoid. Improved clustering accuracy by 51% using a new approach of Deep Clustering for Multi-level Image Segmentation

Machine Learning Engineer | Pragnakalp Techlabs, Ahmedabad, India — Aug 2019 – Jun 2021

- Developed and maintained Docsaar- A document parsing website with HTML- CSS-JS, Flask, MySQL, AWS EC2, Lambda, S3
- Boosted the Named Entity Recognition model for biomedical entities by an average of 5% F1 Score per entity and improved the multi-label Relation Classification model by 4.2% Micro F1 Score. Achieved 78.4% Macro F1 Score on a three-class Sentiment Analysis model and 55% precision on LDA-based topic model for Life Sciences forums and blogs data
- Created "Song playlist Generation system using OpenCV and Deep Learning(CNN)" and published it (Research-Paper)

TECHNICAL SKILLS

Programming languages:	Python, R, Java, JavaScript, MATLAB, C, SQL
Frameworks and Packages:	TensorFlow, PyTorch, Hugging Face (Transformers), Django, Flask, NLTK, PySpark, MongoDB
Tools and Technology:	Image Processing, ML-Ops, AWS, Azure, Databricks, PowerBI, Tableau, SAP, Git
Databases:	Microsoft SQL Server, MySQL, Oracle, PostgreSQL, HBase.

RELEVANT PROJECTS

- **Potential Client Profiling and segmentation using Large Language Models (Westlake Chemicals, Aug-Dec 2023):** Creating 3 different types of customer profiles using LLM by using Naïve LLM, Retrieval-Augmented Generation (RAG) architecture, and Knowledge Graph, which is used to identify potential clients using K-means, DB-Scan, Neural networks.
- **Hourly energy demand/load forecast:** Setup PySpark streaming ETL pipeline to fetch real-time energy demand. Utilized Spark MLlib and RDDs to apply transformations. Carried out statistical tests (F-tests, ADF, ACF, PACF, Granger). SARIMAX outperformed some deep learning models (Facebook Prophet, and GRU) by statistically significant margins (R^2: 0.83)
- **Image Caption Generator:** Developed image captioning using several state-of-the-art Convolutional Neural Networks(CNN) architectures using TensorFlow. Worked with various architectures including ResNet50, InceptionResNetv2, and DenseNet201. Achieved the highest accuracy of 87% with ResNet50 architecture.
- **Relation Extraction:** Built sentence entities relation extraction model. Implemented feature fusion attention-based architecture with CNN to enhance accuracy compare to (Bi-LSTM, CNNs and BERT). Improved F1 Score per entity by 4.2% and Macro F1-Score by 3.6%.

EDUCATION

MS in Data Science | Michigan State University — Sep 2022 – May 2024

Statistical modeling, Data mining, Deep Learning in Finance, Big Data Analytics, Machine Learning, Natural Language Processing, Image Processing, Cloud Computing

BS in Computer Engineering | Pandit Deendayal Energy University — Jul 2018 – Jul 2022

Data Structures and Algorithms, DBMS, Web Development, Software Development, Object-Oriented Programming