KEVIN PATEL

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

Experienced Data Scientist with a strong background in Generative AI, machine learning, data analysis, and software development. Proficient in business insights Analysis, data-driven decision-making, and developing scalable applications. Skilled in gathering, cleaning, and analyzing large datasets, building predictive models, and leveraging statistical analysis for complex business problems. Capable of integrating machine learning solutions into production environments, with a track record of automating data workflows and deploying AI models that improve business outcomes. Expertise in cloud computing, MLOps, and DevOps practices.

WORK EXPERIENCE

Software Developer | Global Mobility Service, IL, USA — Aug 2024 – Current

- Analyzed business data (e.g., revenue, market trends) to derive actionable insights, optimizing sales strategies and customer engagement.
- Developed a Chrome Extension to automate resume generation, showcasing technical proficiency in delivering custom tools to improve operational efficiency.
- Collaborated with the Product Owner to align project deliverables with business needs, refining product requirements in an agile environment.
- Created a tool using Bash and Python to automate the provisioning of Linux servers, streamlining the deployment process for cloud environments.
- Designed data models and workflows to automate data preparation, enhancing decision-making through predictive analytics and machine learning.

Software and Data Expert | VERN.ai, MI, USA — May 2024 – Oct 2024

- Working on integrating the VERN emotion detection model with the conversation bot to observe user sentiment over the period and re-route their flow of conversation according to it. Engineered flow to collect user data and store it at backend.
- Fine-tuned LLMs like LLaMA to personalize user interactions, increasing user satisfaction through adaptive responses.
- Designed and implemented a real-time data ingestion system using AWS Kinesis Data Streams. Developed AWS Lambda functions to process and transform streaming data. Stored processed data in Amazon S3 and loaded it into Amazon Redshift for analysis. Created ETL jobs with AWS Glue to transform and catalog data. Built interactive dashboards and visualizations with AWS QuickSight to provide insights from the data.

Michigan State University | MI, USA

Data Scientist Capstone Project | Siemens (Michigan State University), USA — Jan 2024 – Apr 2024

- Aimed at enhancing design pattern parameter predictions for engineering teams in multi-objective optimization problems
- Worked 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

Data Scientist Capstone Project | Westlake Chemicals (Michigan State University), USA — Aug 2023 – Jan 2024

- Constructed Potential Client Profiling and segmentation pipeline using Large Language Models.
- 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.
- Applied Retrieval-Augmented Generation (RAG) architecture to create a knowledge graph for customer insights, improving the efficiency of client targeting.

Research Assistant | Michigan State University, MI, USA — 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

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

- Developed AI-powered weather forecasting models using LSTM, GRU, and ARIMA, increasing forecast accuracy by 20%.
- 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 (intern)|Pragnakalp Techlabs, India — Aug 2020 – 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 & Frontend:	Python, R, Java, JavaScript, MATLAB, C, SQL, Angular, ReactJS, Flask, Node.JS m Django
Frameworks and Packages:	TensorFlow, PyTorch, Hugging Face (Transformers), Django, Flask, NLTK, PySpark, MongoDB
Tools and Technology:	LLM, Natural Language Processing, ML-Ops, AWS, Azure, Databricks, Git, CI/CD, Docker
Databases:	Microsoft SQL Server, MySQL, Oracle, PostgreSQL, HBase.
APIs :	RESTful API Design, LLM integration, Microservices Architecture.

RELEVANT PROJECTS

- **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 & BERT). Improved F1 Score per entity by 4.2% and Macro-Score by 3.6%.
- **Cricket World-Cup Data Analysis**: Created Power BI report to identify top players for a cricket team on the scrapped data from espncricinfo with Brightdata website tool, perform ETL on data with pandas and SQL and evaluate various performance metrics for players. This can be helpful to team-management to select team and it reduce time by 5-6 hours per match. [link]
- **Hospital Data Analysis** : Perform data cleaning, data engineering as well as missing value imputation before performing EDA. Developed web-dashboard using Streamlit and deployed it on Heroku. Derived conclusive story and business insights using statistical tests like hypothesis and distribution analysis of data and built predictive model to predict approximate stay length of new patient using regression analysis.
- **Natural Language Query Interface with GPT-3 and MySQL**: Designed and implemented a Natural Language Interface to interact with a MySQL database using OpenAI's GPT-3 API. Leveraged the text-davinci-003 model of GPT-3 to translate natural language queries into SQL statements. Post data retrieval utilized the text-DaVinci model from GPT-3 to craft a user-friendly response based on the SQL query and its results.

EDUCATION

- **MS in Data Science**

  Michigan State University | GPA: 4.0/4.0

  Coursework: Data Mining, SQL, Statistics and Probability, Machine Learning, Deep Learning, Software Development, Natural Language Processing (NLP)

- **Bachelor of Technology in Computer Engineering**

  Pandit Deendayal Energy University | GPA: 9.59/10.0

  Coursework: OOP in Java, Data Structures, DBMS, Computer Networks, Database Management Systems, Design and Analysis of Algorithms, Software design, Cloud computing.