KEVIN PATEL

Phone: (517) 505-9591 Kevin18patel@gmail.com
patelk69@msu.edu Michigan, USA

- Results-oriented AI/ML developer with over 4 years of experience designing, developing, and deploying machine learning models and Artificial Intelligence/ Gen-AI solutions.
- Proven track record of delivering impactful projects, strong expertise in data preprocessing, feature engineering, model selection, and hyperparameter tuning.
- Skilled in Python programming and proficient in popular ML frameworks.
- Excellent problem-solving skills and ability to work collaboratively on complex projects.
- Experience creating visually compelling data visualizations using tools like Tableau, Power BI, and Matplotlib to communicate complex findings. Skilled in data collection, cleaning, and analysis, with a strong ability to communicate complex findings to both technical and non-technical stakeholders.
- Proficiency in statistics, mathematical models, and AWS, Azure, and Databricks tools.

Portfolio: https://kevin200010.github.io/Kevin-Patel.github.io/
LinkedIn: https://www.linkedin.com/in/kevin-patel-50b18a195/
GitHub: https://github.com/kevin200010
Blog Channel: https://medium.com/@kevin18patel
YouTube Channel: **Pypower Project :** https://www.youtube.com/c/PyPowerProjects (co-founder)

Education

M.S. Michigan State University, USA Sept 2022 – April 2024
Master's in Data Science
Subjects : Introduction to Data Science, Mathematical Statistics for Data Science, Big Data Analysis, Data Mining, Applied Machine Learning, Statistical modeling, Data mining, Deep Learning in Finance, Big Data Analytics, Machine Learning, Natural Language Processing, Image Processing, Cloud Computing
GPA : 4/4

B.S. Pandit Deendayal Energy University, India Aug 2018 -July 2022
Bachelor of Computer Engineering
Subjects : Design Analysis and Algorithm, Database Management System, Object Oriented Programming, Web Technologies, Artificial Intelligence, Operating System, Compiler Design
GPA : 9.59 /10

Publications

"*Song playlist generator system based on Facial Expression and Song Mood*"

• Presented in 1st International Conference of Artificial Intelligence and Machine Vision AIMV-21,
• link : https://ieeexplore.ieee.org/document/9670976
• Research paper is based on emotion detection and song mood prediction using different ML and DL Algorithms.

Research/Work Experience

Data Scientist Capstone
Siemens, (remote) USA Jan 2024 – Apr 2024
Mentor : Yashesh Dhebar

- 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, USA May 2023 – Apr 2024
Menager : Jianrong Wang

- 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 Capstone (poster)
Westlake Chemicals, (remote) USA Aug 2023 – Dec 2023
Mentor : Yang Jian

- Focused on finding clusters of similar content websites using LLM models.
- Created a retrieval-augmented generation (RAG) architecture.
- Generated embeddings for large documents using GPT 3.5 Turbo.
- Fine-tuned a summary generator model to enhance RAG accuracy and overall clustering accuracy.
- Engineered a Knowledge Graph (KG) from scrapped text and built a bot for question-answering using Langchain GraphCypherQAChain.

AI/ML (Gen AI) Intern
VERN.AI , East Lansing, USA May 2023 – Aug 2023
Menager : Craig Tucker

- Developed a user-centric offline Question-Answer bot using the **Vicuna(7B) model** and **Langchain**. Leveraged **sentence transformers all-MiniLM-L6-v2** to generate embeddings for efficient searching within **Vector DB** and ensuring targeted information retrieval using semantic search and after summarizing using Vicuna-7B it further processed by **Langchain** for accurate question-answer sequences with managing context within the session.
- 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

Research Data Scientist
India Meteorological Department, Bhopal-India Jul 2021 – Aug 2022
Advisor : Ved Prakash Singh

- EDA, Preprocessing, data cleaning and Feature engineering of Meteorological observation such as Radar data and RSRW data
- 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.
- Build model to predict impactful low-altitude wind speed accurately, that provide prediction to overcome the issue of cone of silence for standard radar observation and deploy it on AWS SageMaker.
- Worked on Deep Clustering for cloud data clustering
- Build Unsupervised Multi-level Clustering for Segmentation of Remote Sensing Imageries using DB-Scan and K-means with adaptive K.
- The unsupervised algorithm which can perform the multi-level clustering of images on the basis of different disjoint feature-sets. Then it merges the different clusters by validating similarity conditions.
- The results of clusters generated at each level were evaluated through DB index, Dunn index.
- Work of Deep Clustering has been published in white paper.

Machine Learning Engineer
Pragnakalp Techlabs, Ahmedabad-India Aug 2019 – Jun 2021
Mentor : Mitul Patel

- It is NLP based product provider start-up, which use Working on appointment booking system using Dialogflow platform
- Developed and maintained Docsaar- A document parsing website with HTML- CSS-JS, Flask, MySQL, AWS EC2, Lambda, S3 .
- Hand gesture digit-detection system using CNN using self-generated dataset by detecting ROI for each data.
- Worked on Chatbot Building using GCP integrating with the Document summarization functionality using NLP concept of Abstractive and Extractive Summarization with the help of cloud function and cloud storage.
- 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)

Projects

Resume Personalization web app & Extension :

- Built and implemented website and Chrome-Extension for providing customized resumes for job for LinkedIn Jobs. Integrated with LinkedIn using Auth0 authentication for securely accessing LinkedIn user profile data such as work experience, education, and skills.
- Leverages client-server architecture using HTTP cycle to design the Web extension. Used Flask, GPT-3.5 Turbo API to generate resume from the user profile at a server and JavaScript to fetch data from page and python to extract information from them.

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. The entire backend was seamlessly integrated with a Streamlit-based user interface for enhanced user interaction.

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, clean and transform data with pandas and evaluate various performance metrics for players.
- Used the resulting Power BI Dashboard to select players for various categories (Fast-Bowler, all-rounder, openers/finisher Batsman) and ultimately choose the best 11 for game. This can be used by team-management to select team and it reduce time by 5-6 hour per match.

Hospital Stay length Prediction

- Build web-dashboard using PowerBI as well as Streamlit and deployed on Heroku to come-up with conclusive story and business insights and build predictive model to predict approximate stay length of new patient.
- Perform data cleaning, data engineering as well as missing value imputation before performing EDA.

Hourly energy demand/load forecast

- Setup PySpark streaming ETL pipeline to fetch real-time energy demand from ISO New England SAP S/4 HANA server. Utilized Spark MLlib and RDDs to apply transformations, carried out 5 statistical tests (F-tests, ADF, ACF, PACF, Granger)
- Determined orders for Seasonal ARIMA with exogeneous inputs (fuel prices, weather conditions, transmission costs) SARIMAX outperformed Holt-winters, Negative binomial GLM, and VAR by statistically significant margin (R-squared: 0.83)

GroceryHub app :

- Developed a robust multi-user grocery hub web application utilizing React.js, Node.js, and CSS. Implemented distinct user roles (Customer, Salesman, Manager) with functionalities for order placement, sales analytics, and product management using MYSQL.
- Implemented website using RESTFUL APIs following MVC architecture pattern to ensure separation of concerns.

Virtual Mouse

• The main idea behind this project is to enable the curser moment with hand fingers gestures. Used OpenCV for image processing and apply different filters to locate fingers.

Bharat HandiCraft Website (E-commerce)

- Developed a website for Handicraft Item dealer to take order for different products.
- Used ReactJS framework to develop the frontend UI, NodeJS to handle backend queries and Used MongoDB for database.

Disk-Scheduling and shortest path-finding Algorithm Visualizer

• website to visualize the working of disk scheduling algorithm like FCFS, LOOK , C-LOOK, SCAN & C-SCAN.

Professional Achievements

- Finalist of SSIP Hackathon 2020 (State-Level hackathon (TEAM)) :
  Problem Statement :Garbage-collection system automation using AI
- Winner of HACKFACT in 2020 (PDEU University In-House Hackathons(TEAM)):
  Restaurant management system
- Hacktober-Fest Completer (Open-Source Contribution (SOLO)):
  Select in the top 10000 from 169000

Professional Certificates And Courses

- **Joy of Computing with Python (NPTEL) , IIT Madras**
- **SQL, Stanford-Online**
- **What is Data-Science?(IBM-Coursera)**

Computer Skills

Programming Languages & Frameworks :	Python, R, C++, C, JavaScript
Libraries:	Python Libraries (OpenCV, OpenAI gym, TensorFlow, Keras, Scikit-Learn, Pandas, PyTorch, NumPy, Flask API), JSON, TypeScript, JQuery, Bootstrap.
Web Development Languages and Tools:	HTML5, CSS, Nodejs, VScode, Git, Github, ReactJs, .NET Core, ASP.Net, VB, Web API/Web Service, SQL Server, T-SQL.
Data Science Skills	Data Engineering, Predictive Modeling, Power BI, Data-Visualization, Time-Series Analysis, Decision Analysis, Big Data Analysis (Hadoop)
Probability and Statistics:	Gaussian mixture models, hidden Markov model, Baye's Theorem.
Cloud/Databases:	RStudio, GoogleColab, Microsoft Azure, Amazon WebServices (EC2, S3, SageMaker), Docker, Oracle, Git, Apache Spark, ETL tools.
Machine Learning and Deep learning :	Linear-logistic regression, Clustering, SVM, PCA, Random Forest, Boosting, Lasso, Ridge, CNN, RNN, TensorFlow, Keras.
Databases:	Microsoft SQL Server, MySQL, Oracle, PostgreSQL, HBase.
Analytics:	Jupyter Notebook, R Studio, MS Excel, SSIS, ETL, MS Access, Tableau, Power BI, Google Analytics.