Kshitij Patel

Los Angeles, CA • patelksh@usc.edu • 774-262-8275 • linkedin.com/in/kshitij-patel-20 • github.com/kapatel20

EDUCATION

Master of Science, Computer Science   **Aug 2022-May 2024**
University of Southern California, Los Angeles, California   **GPA: 3.83**
Course Producer: DSCI 552 : Machine Learning for Data Science

Bachelor of Technology, Computer Engineering   **Aug 2018-May 2022**
Pandit Deendayal Energy University, Gandhinagar, India   **GPA: 4.00**

TECHNICAL SKILLS

Programming: Python, R, SQL, C++, C#(.Net Framework), C, Java, JavaScript ES6, PHP
Data Science/ML: Statistical Modelling/Analysis, ETL Pipelines, TensorFlow, PyTorch, Transformers, NLTK, PySpark, Scikit-learn, NumPy, Pandas, Matplotlib, Regression, Classification, Decision Tree, SVM, Clustering, PCA, Transfer Learning
Tools/Database: GIT, Tableau, AWS(S3, Kinesis), GCP, Jenkins, Docker, Elastic Search, Kibana, Hadoop, MapReduce, Spark

EXPERIENCE

Software Development Intern, *Ria Money Transfer*, Buena Park, USA   **Jun 2023-Aug 2023**
- Developed C#(.NET Framework) based Application to streamline data extraction, processing, and server ingestion from internal **10000+** customer files using APIs. Saved **3-5 hours of manual data handling time per week** and improved operational efficiency.
- Leveraged elastic-search and kibana to perform analysis of log files. Preprocessed time-series data and trained LSTM-based model to forecast trends of API failures. Achieved **MSE of 4.287** and contributed to enhance decision-making capabilities.
- Demonstrated adeptness in Large Language Models(LLMs) and leveraged HuggingFace Transformers library to fine-tune GPT-2, RoBERTa and ALBERT models to develop in-house question-answering system for company's API documentations.

Machine Learning Intern, *Silver Touch Technologies Ltd.*, Ahmedabad, India   **Jan 2022-Apr 2022**
- Employed YouTube APIs to extract meta-data about **40000** videos and formulated advanced content recommendation engines utilizing **BERT, FastText, and TF-IDF Vectorizer** to provide personalized suggestions, yielding **Intra-Similarity score of 0.7233**.
- Productionalized recommendation system using **Flask-based REST API** for deployment and maintained dataset.

Undergraduate Research Assistant, *Pandit Deendayal Energy University*, Gandhinagar, India   **Aug 2021-Dec 2021**
- Conducted research under Prof. Samir Patel to identify abnormalities in musculoskeletal radiographs. Leveraged **transfer learning and deep neural networks** on MURA dataset to detect fractures in seven different body parts like Elbow, Fingers, etc.
- Applied Edge Detection and Image Sharpening Techniques to enhance base model **accuracy by 5%**. Achieved average **Precision value of 84.55%** and outperformed **Stanford ML Group's** performance in case of Fingers **by 11%**.

Data science Intern, *Silver Touch Technologies Ltd.*, Ahmedabad, India   **May 2021-Sep 2021**
- Conducted exploratory data analysis of about **20000+ schools** in team. Adapted Pandas for data cleaning and preprocessing and applied **statistical analysis** and employed **data visualization** libraries to translate data insights into graphical representations

PROJECTS

Generating SQL Queries from Natural Language — Code   **Jan 2023-May 2023**
Python, PyTorch, Natural Language Processing, Semantic Parsing
- Implemented BiLSTM Seq2Vec models in PyTorch based on **ACL publication** for tasks including selecting aggregation functions, selecting operators, and selecting columns in SELECT and WHERE clauses.
- Employed **attention mechanism** between columns and questions to enhance the baseline accuracy by more **than 10%**.

Time Series Classification for Human Activity Recognition — Code   **Jan 2023-May 2023**
Python, Scikit-Learn, Statsmodels, Statistical Analysis, Time-Series Analysis
- Conducted time-series analysis for seven different human activities, extracting seven time-domain features for each activity.
- Utilized cross-validation and backward selection methods to find statistically significant features and trained classification models to classify activity, achieving an **accuracy of 94.73%** using **Multinomial Naive Bayes algorithm**.

Sentiment Analysis of Amazon Customer Review Dataset — Code   **Jan 2023-May 2023**
Python, PyTorch, Scikit-Learn, NLTK, Gensim, NLP, Sentiment Analysis, Machine Learning, RNN, LSTM, GRU
- Preprocessed data, extracted TF-IDF word embeddings and experimented models including **SVM, RNN, LSTM, and GRU** to perform **multiclass classification**. Utilized **word2vec-google-news-300** for word embeddings to increase **base accuracy by 5%**.

Game Playing Agent for Go (Top 10 ranking among 240 peers) — Code   **Aug 2022-Dec 2022**
C++, Artificial Intelligence
- Implemented AI Agent for **5x5 Go board game**. Designed custom heuristic to find best move for depth of 7 plys using **Alpha-Beta Pruning** for **Minimax algorithm**. Defeated Random, Greedy, Aggressive, Alpha Beta and Q-learning agents in **over 95% of games**.