Sumit Mishra

LinkedIn | Github

GET IN CONTACT

Mobile: +91-7400255701

Email: sm27598@gmail.com

PERSONAL DETAILS

- Total Experience — 5 Years 2 Months
- Current Location — Mumbai
- Date of Birth — May 27, 1998
- Gender — Male
- Marital Status — Single / Unmarried

CORE SKILLS

- Python
- Natural Language Processing
- Machine Learning
- Large Language Model (LLM)
- Deep Learning
- Neural Networks
- Pytorch
- Flask
- Apache Airflow
- SQL
- Scikit-Learn
- Pandas
- Numpy
- Spacy
- Matplotlib

EXPOSURE TO

- Apache Kafka
- PySpark
- Keras
- Exploratory Data Analysis
- Prompt Engineering
- OCR
- HTML
- CSS
- Javascript
- JQuery
- Core PHP
- ASP.NET

PROFILE SUMMARY

Skilled AI/ML engineer with overall 5+ years of experience, including 3+ years of experience developing NLP, computer vision, and data pipeline solutions using Python, Spacy, Kafka, Airflow, Numpy, Pandas, and SQL. Core strengths in transformer models such as BERT & Flan-T5, OpenCV, PyTorch, LangChain, and building web applications using Flask to support annotation efforts. Collaborated cross- functionally with USA-based professors in AI, engineers, and marketers across various projects.

WORK EXPERIENCE

Jan 2022 to Present

Research Assistant (Natural Language Processing & Machine Learning) at University of Utah

- Developed various NLP-related products, including a Semantic Similarity Matcher, Rule-Based POS tagging models, and text summarization and classification models. Also, created an Optical Character Recognition (OCR) project for Ad Analysis.
- Below are details of a few key projects I undertook during my tenure:

Semantic Sentence Similarity Matcher
- Created the 'Semantic Sentence Similarity Matcher' by utilizing cutting-edge NLP models, including BERT, BART,and T5.
- As part of this project, tried various embeddings, fine-tuned the classification layer, and adjusted all the layers of the model to obtain the best-optimized model for sentence similarity.
- Also trained a regression model for sentence similarity to predict the probability of sentences being similar using the aforementioned transformer-based models.

Summarization Tagging
- Created the research project using Spacy's rule-based techniques to tag and summarize sentences in articles. Utilized POS tagging, token matcher, phrase matcher, and dependency matcher.
- Based on labeled sentences from articles, identified patterns to create a Spacy rule-based system for categorizing sentences.
- Additionally, designed, developed, and deployed a web application using Python and Flask to facilitate the testing and validation of the rule.

- C#

LANGUAGES KNOWN

- English
- Hindi
- Marathi

COURSES & CERTIFICATIONS

- Applied Machine Learning Course
- Python For Machine Learning, Statistics for Data Science & Machine Learning, Data Visualization, Etc.

AWARDS

- **1st** Rank in 3P (Project-Prototype-Presentation) Evaluation Event, awarded by B.Sc. (I.T.) Association.
- **3rd** Rank (81.6%) in Semester 5 University Examination for the Year 2017-18, awarded by V.E.S. College.
- **1st** Rank (145/150) in the Advanced Java University Examination for the Year 2017-18, awarded by V.E.S. College.
- **1st** Rank in Quiriosity Event of Vihaan'18, an Inter-Collegiate Technical Fest.
- **2nd** Rank in Iotics (Internet of Things) Event Of Vihaan'18, An Inter-Collegiate Technical Fest.

VOLUNTEER & EXTRA-CURRICULAR ACTIVITIES

- Organized an event on cyber-security for college students.
- Volunteered for organizing a technical fest of the B.Sc. (I.T.) department of V.E.S. college.
- Helped and Participated in Annual Extension Activities (Lifelong learning & Extension) related to Information Technology, Industry Orientation, Entrepreneurship & Career Guidance, etc.
- Engaged in various inter-collegiate technical competitions and disaster awareness programs, showcasing a commitment to learning and community service.

- To learn more about the projects undertaken, please visit this link: https://bit.ly/3Zbs1HY

May 2021 to May 2022

Data Engineer at mentorstudents.org

- Designed and implemented automated data pipelines using Apache Kafka in Docker and Apache Airflow for efficient data handling and processing.
- Developed ETL projects to create an efficient data pipeline for NLP and ML.
- Below are details of a few key projects I undertook during my tenure:

StreamFlow: Automated Daily Data Pipeline
- o Designed and implemented an automated data pipeline fordaily data streaming and processing using Apache Kafka and Apache Airflow within a Docker environment.
- o Created Directed Acyclic Graphs (DAGs) to automate the data ingestion and processing, enabling us to send daily insights to our clients.

Social Media Data Analysis and Classification
- o Designed and developed a robust data pipeline for socialmedia sentiment analysis and zero-shot classification to gather insights into what customers are most frequently discussing about their products.
- o Utilized a RESTful API to collect data from social media andemployed a pre-trained zero-shot classification model to analyze the sentiments of users in relation to the client's product.
- o To streamline this process, orchestrated it within a datapipeline using Apache Kafka and Apache Airflow and created a dashboard to provide visual insights.

- To learn more about the projects undertaken, please visit this link: https://bit.ly/3Zbs1HY

Mar 2021 to May 2021

Data Science and Business Analytics Intern at The Sparks Foundation

- Wrote code to collect, process, and analyze data from both internal and external sources, enhancing data-driven decision-making.
- Developed and fine-tuned machine learning models using Python and scikit-learn, improving predictive capabilities.
- Successfully completed various projects in Machine Learning, NLP, Deep Learning, and Computer Vision.
- Leveraged skills in Keras for deep learning projects, enhancing model accuracy and performance.
- Gained hands-on experience in data analysis and natural language processing (NLP), contributing to data-driven insights.

Sep 2019 to Mar 2021

Software Developer at TalkValley LLC

- Developed responsive web applications, ensuring user-friendly interfaces and optimal functionality.
- Collaborated cross-functionally to gather requirements and create effective web solutions.
- Integrated third-party APIs to enhance application functionality, improving user experience.
- Conducted rigorous testing and debugging to maintain seamless application performance.
- Successfully maintained and updated existing web applications, staying current with emerging technologies.
- Ensured cross-browser compatibility, delivering consistent user experiences across platforms.
- Contributed to code quality through active participation in code reviews and provided technical support for issue resolution.

Aug 2018 to Sep 2019

Programmer at Tata Consultancy Services Ltd.

- Applied operating system and program product fixes regularly, ensuring the stability and reliability of systems.
- Ensured strict adherence to process conformity and consistently achieved SLAs (Service Level Agreements).
- Managed network devices, servers, batches, and jobs effectively during transitions to prevent irrelevant and erroneous issues.
- Collaborated with various mainframe services to address specific requests and optimize system performance.
- Liaised with software vendors to report, track, and efficiently resolve issues related to software products, ensuring seamless operations.
- Coordinated external dependencies with neighboring teams and subject matter experts (SMEs), facilitating smooth project execution.

EDUCATION

Post Graduation

M.Sc (Data Science)
Chandigarh University
Jan 2021 to Present

Graduation

B.Sc (Information Technology)
University of Mumbai
Jun 2015 to May 2018