

Sumit Mishra (He/Him/His)

Chembur, Mumbai, Maharashtra 400074, India

sm27598@gmail.com

(+91) 7400255701, (+91) 8689907119

[LinkedIn](#) | [Website](#) | [Github](#)

EDUCATION

Chandigarh University

Chandigarh, Punjab, India (Jan 2023 - Present)

Master of Science in Data Science

- Relevant Coursework: Applied Probability and Statistics, SQL, Communication and Soft Skills, Python, Advanced Database Management System.

University of Mumbai

Mumbai, Maharashtra, India (June 2015 - May 2018)

Bachelor of Science in Information Technology

- Relevant Coursework: Applied Mathematics, Web Technologies, Database Management System, Logic and Discrete Mathematics, Computer Graphics, Advanced SQL, Java, Advanced Java.
- Capstone Project: "Theft Prevention Using PIR Sensor (IoT project)"
- Grade: A

RESEARCH EXPERIENCE

University of Utah

Salt Lake City, Utah, USA (Jan 2022 - June 2023)

Research Assistant (Remote)

- Engaged in research endeavors aimed at advancing text summarization techniques and harnessing advanced machine learning models to extend the horizons of knowledge within the field. Conducted thorough investigations into the effectiveness of the PRIMER model for multi-document summarization, synthesizing insights from the PRIMER model and relevant scholarly literature to inform experimental initiatives.
- Demonstrated a comprehensive understanding of language modeling and multi-document summarization through an in-depth review of academic papers. Navigated complex workflows using tools like the Fairseq CLI tool and the PyTorch framework. Translated intricate Fairseq implementations into PyTorch modules, optimizing the implementation process for seamless model integration.
- Contributed to the automation of data collection efforts by developing scripts interfacing with the Google API. Designed a customized Flask application for efficient article preview and manual annotation, significantly accelerating the annotation workflow. The collected and annotated data played a crucial role in model fine-tuning across various experiments.
- Made contributions to the machine learning landscape by refining transformer-based embedding models, laying the groundwork for improvements in heading and section similarity. Executed model fine-tuning tasks involving semantic textual similarity and question categorization, demonstrating proficiency in these domains.
- Facilitated complex system deployments using tools such as Apache, MySQL PHPMyAdmin, NGINX, WSGI, and Unicorn. Managed extensive datasets, resulting in optimized model performance. This system played a key role in the research paper titled "Enhancing Prompt Efficiency for Non-native English Speakers via Metadata Prompting: a Field Study," rigorously examining metadata prompting's efficacy in improving prompt efficiency for non-native English speakers through a comprehensive field study.
- Contributed to iterative application enhancements, effectively addressed challenges, and refined methodologies through transparent research processes. Maintained documentation practices and organized code repositories, implementing tracking mechanisms for effective project management.

PROFESSIONAL EXPERIENCE

TalkValley LLC

Mohali, Punjab, India (May 2021 - May 2022)

Data Science Analyst (Remote)

- Executed a series of live projects that showcased expertise in PySpark, Apache Airflow, Kafka, and advanced NLP techniques.
- Collaborated on the creation of an end-to-end Airflow project, integrating sensor-based file checks, MySQL interactions, report generation, and email notifications to stakeholders.
- Implemented a high-performance data pipeline using Kafka, incorporating real-time data ingestion and database backup tasks to ensure data integrity.
- Conducted comprehensive testing and validation of sentiment analysis and zero-shot classification projects, ensuring seamless end-to-end functionality.
- Played a pivotal role in spearheading innovative NLP projects, delving into cutting-edge applications. Did the design and implementation of a sentiment analysis framework leveraging dense and sparse embeddings, providing actionable insights into customer feedback.
- Additionally, contributed to the creation of a novel chatbot using sequence-to-sequence models, enhancing user engagement and interaction on the company's digital platform.

Key Achievements:

- Contributed to the successful execution of impactful live projects, showcasing the ability to deliver practical and effective data solutions.
- Collaborated on the implementation of an advanced Apache Airflow project, demonstrating adaptability and proficiency in mastering new technologies.
- Elevated NLP's impact by driving innovative projects, including sentiment analysis implementation for customer feedback and the deployment of an interactive chatbot using advanced sequence-to-sequence models.
- Actively engaged with the mentorstudents.org platform, fostering a culture of knowledge exchange and community learning.

Skills Demonstrated:

- Data Engineering: PySpark, Apache Airflow, Kafka
- Data Analysis: SQL, Data Profiling, Visualization
- Machine Learning: PySpark ML, Sentiment Analysis, Zero-shot Classification
- NLP Techniques: Dense and Sparse Embedding, RNNs, LSTMs, CNNs, BERT, GANs
- Databases: MySQL, Cassandra
- Documentation: Project Descriptions, Instructions, Workflow Explanations

The Sparks Foundation

Data Science & Business Analytics Intern (Remote)

(Mar 2021 - May 2021)

- Wrote codes to collect, crunch and analyze data from internal and external sources.
- Built and tuned various machine learning models using Python, and scikit-learn.
- Created linkages between various data to enable predictive modeling and trend analysis on a given dataset.
- Built various Machine Learning, NLP, Deep Learning, and Computer Vision-based projects.

TATA Consultancy Services

Programmer

Chennai, Tamilnadu, India (Aug 2018 - Sep 2019)

- Regularly applied operating system and program product fixes.
- Ensured adherence to process conformity and achievement of SLAs.
- Effectively managed network devices, servers, batches, and jobs during transitions to mitigate irrelevant and erroneous issues.
- Collaborated with various mainframe services to address specific requests.
- Liaised with software vendors to report, track, and resolve issues related to software products.
- Coordinated external dependencies with neighboring teams and subject matter experts (SMEs).

ADDITIONAL EXPERIENCE

Chegg India

(Sep 2021 - Present)

Managed Network Expert (Remote)

- Providing detailed step-by-step solutions to student queries in Computer Science, facilitating comprehensive understanding.
- Break down complex concepts into simplified explanations to enhance student learning and comprehension.

Freelance

(May 2016 - Present)

Full Stack Developer

- Developed a range of projects catering to diverse clients, including colleges, students, stores, and academies.
- Created an Auditorium & AV Room Booking System to automate the booking process, replacing manual methods in colleges. The intranet-based system offers real-time slot availability and direct booking capabilities.
- Designed a comprehensive Swimming Academy Website featuring user and admin dashboards. Integrated with Instamojo payment gateway, allowing seamless fee collection. Admin has control over user management, content creation, fee modification, and partnership modules.
- Crafted a dynamic Mind & Body Yoga Class Website, showcasing services, client testimonials, and published blogs. Implemented security measures with Captcha, leveraged Google Analytics for insights, and optimized the site with Google Ads services.

PUBLICATIONS & BLOGS

- Publication: *"Enhancing Prompt Efficiency for Non-native English Speakers via Metadata Prompting: a Field Study"*
 - Submitted at the EMNLP 2023 Conference.
 - Authors: Rohit Aggarwal, Sumit Mishra, Harpreet Singh
 - This paper explores the effectiveness of metadata prompting in improving prompt efficiency for non-native English speakers, as demonstrated through a field study.
- Blog: *"Scene Text Detection, Recognition & Translation"*
 - Published in "The Startup" publication | Jan 2021
 - This blog discusses the techniques and challenges related to scene text detection, recognition, and translation, providing insights into this evolving field.
- Blog: *"Stack-Overflow Search Engine & Question Recommendation"*
 - Personal Blog | Aug 2020
 - The blog delves into the development of a search engine tailored for Stack Overflow, including question recommendation techniques to enhance user experience.
- Blog: *"Exploratory Data Analysis on Corona Virus Data"*
 - Personal Blog | Mar 2020
 - This blog showcases an exploratory data analysis of coronavirus data, offering insights into patterns and trends during the initial stages of the pandemic.

SKILLS

- Programming Languages: Python, Java, C++, JavaScript, SQL
- Machine Learning & NLP: Natural Language Processing, Machine Learning Algorithms, Text Classification, Text Summarization, Language Modeling, Sentiment Analysis, Neural Networks, Deep Learning
- Data Science & Engineering: Data Analysis, Data Preprocessing, Feature Engineering, Data Visualization, Data Cleaning, Exploratory Data Analysis, Database Management Systems
- Tools & Frameworks: JQuery, TensorFlow, PyTorch, Keras, Scikit-Learn, Pandas, NumPy, Matplotlib, Seaborn
- Languages: Professional working proficiency in English, Native or bilingual proficiency in Hindi, Limited working proficiency in Marathi

CERTIFICATIONS

- Applied Machine Learning
 - By Applied AI Course (Feb 2020 - Feb 2021)
 - Details: Completed comprehensive coursework including 30+ assignments and successfully completed two capstone projects focused on NLP and Computer Vision.
- Data Visualization Using Python
- Statistics for Data Science
- Data Science with Python
- Python for Machine Learning
- Statistics for Machine Learning
- Machine Learning with Python
- Machine Learning Foundations
- Introduction to Neural Network & Deep Learning
- Computer Vision Essentials
 - By Great Learning (Mar 2020 - Apr 2021)

AWARDS

- **1st** Prize in 3P (Project-Prototype-Presentation) Evaluation Event, awarded by BSc (I.T.) Association.
- **3rd** Rank (81.6%) in Sem 5 University Examination for the Year 2017-18, awarded by V.E.S. College.
- **1st** Rank (145/150) in 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.
- **1st** Prize in Other Extra-Curricular Events of the Fest such as NIRAYA, an Inter-Collegiate Fest.

VOLUNTEER & EXTRA-CURRICULAR ACTIVITIES

- Organized an event on cyber-security for college students.
- Volunteered for organizing a technical fest of the Bsc(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.