MANDEEP

+1 (236) 862-1674 | mandeepkhokhar1110@gmail.com | Vancouver, BC, Canada | linkedin.com/in/mandeep-0708a2186/ | github.com/M-I-Dx

EDUCATION

University of British Columbia **September 2023 - July 2024**
Master's, Data Science in Computational Linguistics *GPA: 93.4*

- Advanced Natural Language Processing: Focused on transformer models, sequence-to-sequence tasks, and domain-specific language models.
- Machine Learning and Optimization: Explored supervised and unsupervised learning, optimization algorithms, and their applications in large-scale data problems.
- Computational Linguistics: Explored syntax, semantics, and discourse analysis using computational methods.
- Data Visualization and Interpretation: Worked on techniques for visualizing complex data, focusing on interactivity and clarity.
- Master of Data Science - Computational Linguistics International Scholarship

Thapar Institute of Engineering and Technology **August 2019 - July 2023**
Bachelor's, Computer Engineering *GPA: 84.5*

- Data Structures and Algorithms: In-depth study of algorithms, their complexities, and data structure optimization
- Artificial Intelligence and Machine Learning: Introduction to AI, covering fundamental algorithms and machine learning models.
- Database Management Systems: Design and implementation of relational databases, focusing on SQL and NoSQL databases.
- Software Engineering: Software development lifecycle, agile methodologies, and project management.

SKILLS

Skills: Python, R Programming, JavaScript
Machine Learning: Pandas, NumPy, SciPy, PyTorch, TensorFlow, Keras, Hugging Face, Numba, NLTK, SpaCy
Deep Learning: OpenAI API, Large Language Models (LLM), Prompt Engineering, Natural Language Processing (NLP), Transformers, LSTMs, RNNs, Convolutional Neural Networks, Recommender Systems
Visualization: Altair, Plotly, Plotly-Dash, Seaborn, Matplotlib, Looker Studio
Databases: Relational database design, PostgreSQL, MongoDB, Firebase, BigQuery
Others: Git, GitHub, A/B testing, Docker, Weights & Biases (Wandb), Google Cloud Platform, Azure ML, Microsoft Fabric, Linear Programming, HiGHS, CPLEX

PROFESSIONAL EXPERIENCE

Canfor **Vancouver, BC, Canada**
Data Scientist *July 2025 - Present*

- Leading the development of a large-scale optimization program to reduce logistics and shipping costs by translating complex business constraints into mathematical formulations, leveraging linear programming and mixed-integer optimization techniques.
- Designed and implemented solver-agnostic optimization pipelines using SciPy (HiGHS), SCIP, CPLEX, and Gurobi, enabling robust comparison of solution quality, performance, and scalability across solvers, with projected annual cost savings of approximately $1.5M.
- Collaborating closely with operations, supply chain, and business stakeholders to formalize real-world constraints, such as capacity limits, routing rules, and demand priorities, into objective functions and constraint sets suitable for mathematical optimization.
- Developing time series forecasting models to estimate production outputs by acquiring domain-specific knowledge of manufacturing processes and evaluating a wide range of techniques, including Chronos2, TSMixer, XGBoost, and LightGBM.
- Building and validating forecasting pipelines using modern time series libraries such as unit8-darts, conducting rigorous backtesting, model comparison, and error analysis to select models that balance accuracy, robustness, and interpretability.
- Deploying, training, and orchestrating machine learning workflows using Azure Machine Learning and Microsoft Fabric, ensuring scalable experimentation, reproducibility, and integration with enterprise data platforms.

CoffeeSpace **Remote**
Machine Learning Engineer *April 2025 - July 2025*

- Led the enhancement of a people-to-people recommender system by integrating machine learning techniques, with a strong focus on feature engineering using Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG).
- Engineered analytics pipelines to measure user behavior and evaluate recommendation quality, leveraging BigQuery and Looker Studio for large-scale data analysis and detailed metric reporting.
- Developed and deployed cloud-native solutions using Python, JavaScript, and Firebase, and implemented scalable services on Google Cloud Platform (GCP) via Cloud Run.
- Designed and integrated vector database solutions, including querying the Qdrant vector database, to improve similarity search capabilities within recommendation systems, significantly enhancing personalization and retrieval efficiency.
- Optimized recommendation and analytics systems by building robust feature extraction processes, training workflows, and model deployment strategies, ensuring smooth end-to-end operations across cloud environments.

Betterdata **Remote**
Machine Learning Engineer *February 2022 - July 2023*

- Spearheaded research in tabular synthetic data generation, implementing cutting-edge generative model architectures that ensure data diversity, fidelity, and privacy.
- Developed and integrated differential privacy into generative models using PyTorch, ensuring compliance with data protection standards.
- Adapted generative model code for versatile deployment, both cloud-based and on-premises, optimizing training and evaluation pipelines using the Numba library for parallel processing.
- Engineered an evaluation system measuring relevance, accuracy, and privacy of generated data, utilizing TensorFlow for model prototyping, Scikit-learn for statistical analysis, and PyTorch for implementing deep learning-based metrics.

Thapar University **India**
Research Assistant *January 2020 - July 2022*

- Researched convolutional neural networks and the effect of hyperparameter tuning and architectures on explainable AI techniques like XAI: LIME and SHAP, using libraries such as TensorFlow and Keras.
- Explored the application of generative adversarial networks in enhancing deep learning detection systems by generating synthetic COVID-19 CT and MRI scans utilizing PyTorch and OpenCV.
- Studied the efficacy of machine learning methods in forecasting the social media traction of various topics, employing libraries such as Numpy, Pandas, and SQL for data analysis and manipulation.

SpaceML **Remote**
AI Researcher *February 2021 - February 2022*

- Collaborated with NASA's IMPACT team to develop an AI system utilizing Self-Supervised and Active Learning techniques, enabling swift identification of pertinent satellite imagery from petabyte-scale datasets.
- Spearheaded the creation of a scalable Active Learning pipeline, integrating multiple cloud scalability layers for cost efficiency using AWS and Google Cloud Platform (GCP).
- Achieved a drastic reduction in manual image labelling time, decreasing it from 7000 hours to 52 minutes for a five million-image climate dataset, utilizing libraries such as TensorFlow and PyTorch.
- Open-sourced a high-quality package on GitHub that is already used for labelling in NASA's Phenomenon portal.
- Secured a NASA Science Mission Directorate grant with the proposal ranking in the top 5 out of 79 research initiatives.

PROJECTS

Biomedical Lay Summarization using Large Language Models *March 2024 - June 2024*

- Generated accessible summaries of biomedical research for non-expert audiences, aiming to improve public understanding of complex scientific findings.
- Designed and optimized domain-specific large language models (LLMs) like BioMistral7b, focusing on prompt tuning and dynamic instruction crafting for enhanced zero-shot and few-shot summarization.
- Implemented retrieval-augmented generation (RAG) and representation engineering to improve prompt responses, aligning better with factual accuracy and user needs.
- Developed prompt-based control mechanisms to dynamically adjust model output for better readability, technical relevance, and faithfulness to the original research.
- Achieved state-of-the-art results in biomedical lay summarization, significantly improving the accuracy, clarity, and accessibility of research summaries for public audiences.

Development of an Annotated Corpus for Problem-Solving Explanations *September 2023 - December 2023*

- Designed and implemented advanced prompt engineering techniques to improve problem-solving capabilities in language models.
- Created an annotated corpus based on the AI2 Reasoning Challenge (ARC) dataset to fine-tune large language models (LLMs) for generating structured, step-by-step explanations using the “Tree of Thought” prompting methodology.
- Developed a detailed annotation framework to break down problems into logical sequences, generating step-by-step natural language explanations using the “Tree of Thought” reasoning framework.
- Significantly improved the model’s ability to solve complex problems from the ARC dataset, enhancing reasoning accuracy and transparency.

PUBLICATIONS

- Deep learning-based explainable target classification for synthetic aperture radar images
- SpaceML Worldview Search - Learnings from an AI citizen scientist team building a NoCode Data Curator from Unlabeled Petabyte Scale Imagery
- Smartphone Based Grape Leaf Disease Diagnosis and Remedial System Assisted with Explanations
- Machine Learning Based Explainable Financial Forecasting