

Mohinur Yusufjanova

(970) 732-1952 | mohinur_yusufjanova@mines.edu | Fort Collins, CO 80526

Master's student in Data Science at Colorado School of Mines with research experience in artificial intelligence, machine learning, computer vision, and large-scale data mining. Experienced in developing Python-based data pipelines, applying machine learning techniques to complex datasets, and conducting independent research in academic environments.

Current research focuses on AI-driven information extraction, OCR, and data mining from large collections of historical documents, while additional work includes computer vision, predictive modeling, and scientific computing. Passionate about developing AI systems that support real-world decision-making, scientific discovery, and complex operational challenges.

EDUCATION

Master of Science in Data Science

Expected in May 2027

Colorado School of Mines, Golden, CO

GPA: 3.9

Relevant Coursework: Advanced Data Science, Machine Learning, Advanced Machine Learning, Advanced Statistics I & II, OOP

Bachelor of Science in Business Administration

August 2019 – August 2024

Colorado State University, Fort Collins, CO

GPA: 3.6

TECHNICAL SKILLS

- Artificial Intelligence & Machine Learning: Python, Scikit-Learn, TensorFlow, PyTorch, Neural Networks, Deep Learning, Computer Vision, Classification, Feature Engineering, Model Evaluation, Statistical Learning
- Data Science & Scientific Computing: NumPy, SciPy, Pandas, Jupyter Notebook, Statistical Analysis, Data Mining, Predictive Modeling, Experimental Design
- Data Engineering & Information Extraction: OCR, Information Extraction, ETL Pipelines, Web Scraping, Data Validation, PDF Processing, Dataset Development, Structured Data Generation
- Programming & Software Development: Python, SQL, R, JavaScript, Git, GitHub, Linux, Bash, VS Code
- Research & Reproducibility: Literature Reviews, Scientific Documentation, Technical Writing, Reproducible Research Workflows, Data Pipeline Development

RESEARCH

Research Assistant – AI & Data Mining

Summer 2026 – Present

- Support a research project on the history of economic geology education across U.S. universities
- Develop Python-based workflows to locate, collect, and process historical course catalogs from online archives.
- Apply web scraping, OCR, information extraction, and data mining to build structured research datasets.
- Extract, validate, and organize course information from large collections of archival records for downstream analysis.
- Build reproducible data pipelines, maintain datasets, and document workflows supporting ongoing research efforts.
- Conduct background research and literature reviews on historical academic programs and archival datasets.

PROJECTS

GridGPT: Retrieval-Augmented AI Assistant for Power System Operations

May 2026

- Developed a retrieval-augmented generation (RAG) system that enables natural language querying of power system operations, grid modernization, and distributed energy resource documentation.
- Built an end-to-end AI pipeline using Python, LangChain, FAISS, and transformer-based language models to ingest, process, and retrieve information from large collections of technical reports and research publications.
- Implemented semantic search, vector embeddings, document chunking, and conversational question-answering workflows to support knowledge discovery and decision support tasks.
- Designed an interactive web application for exploring engineering documents and evaluating AI-generated responses across domain-specific technical questions.
- Evaluated retrieval performance and response quality using benchmark questions derived from publicly available energy systems research literature.

University Course Catalog Mining Pipeline

June 2026

- Developed an end-to-end Python pipeline for discovering, processing, and analyzing large collections of historical university course catalogs from online archives.
- Automated document acquisition, OCR, information extraction, entity recognition, and data validation workflows to transform unstructured records into structured research datasets.
- Designed scalable data processing pipelines capable of handling thousands of pages of archival documents while maintaining data quality and reproducibility.
- Applied natural language processing and AI-assisted extraction techniques to identify, organize, and standardize course information across heterogeneous document sources.
- Built reproducible research workflows supporting large-scale analysis, dataset generation, and downstream machine learning applications.

Solar Energy Site Suitability Analysis

April 2026

- Developed a data-driven decision support framework to identify candidate locations for utility-scale solar energy deployment across Colorado.
- Integrated and processed large geospatial datasets related to land use, terrain characteristics, environmental constraints, solar resource availability, and transmission infrastructure.
- Applied multi-criteria analysis, spatial modeling techniques, and quantitative scoring methods to evaluate renewable energy deployment opportunities.
- Built automated Python workflows for large-scale data processing, feature generation, and suitability assessment.
- Generated actionable insights to support energy planning, infrastructure development, and renewable energy decision-making.

Image Enhancement for Monocular SLAM

Ongoing

- Developing a data-driven decision support framework to identify candidate locations for utility-scale solar energy deployment across Colorado.
- Integrating and processing large geospatial datasets related to land use, terrain characteristics, environmental constraints, solar resource availability, and transmission infrastructure.
- Applying multi-criteria analysis, spatial modeling techniques, and quantitative scoring methods to evaluate renewable energy deployment opportunities.
- Building automated Python workflows for large-scale data processing, feature generation, and suitability assessment.
- Generating actionable insights to support energy planning, infrastructure development, and renewable energy decision-making.

EXPERIENCE

Graduate Teaching Assistant - Advanced Data Science / Computer Science

January 2026 – May 2026

Courses: DSCI503 (Advanced Data Science), DSCI403/CSCI413 (Computer Science)

- Review and grade graduate-level projects focused on data analysis, machine learning, and statistical modeling.
- Provide feedback on data preparation, model evaluation, and analytical approaches used in student projects.
- Help students work through technical challenges and better understand data science concepts.
- Explain technical topics in a clear and approachable way during office hours and student interactions.