Mihir Ojha

- *Email:* mihir.ojhaa@gmail.com • *Portfolio:* https://mihirojha.github.io/ • *LinkedIn:* Mihir Ojha

EDUCATION

- **Simon Fraser University** — June, 2022
  Bachelor of Science - Data Science — *Burnaby, BC*
    - Relevant Coursework: Operating Systems, Data Structures, and Algorithms, Statistics, Machine Learning, Database Management

SKILLS

- **Languages**: Python, R, SQL
- **Machine Learning**: Supervised Learning (Regression, Decision Trees, Random Forest, Gradient Boosting, XGBoost), Unsupervised Learning (k-means Clustering, Dimensionality Reduction, PCA), Neural Networks (ANN, CNN), Deep Learning, Hyperparameter Tuning
- **Frameworks**: Scikit-learn, NLTK, SpaCy, TensorFlow, Keras
- **Data Analysis and Modeling**: Pandas, Numpy, SciPy, Data Wrangling, Data Cleansing, Data Preprocessing, EDA
- **Data Visualisation**: Python's Matplotlib, Seaborn, Plotly, Microsoft Power BI and Tableau
- **Extra**: Git version-control, GitHub, HTML, CSS, Agile, Scrum, Linux, Command Line Interface, Heroku, Railway, Streamlit

EXPERIENCE

- **Jain University** — Remote
  Machine Learning Research Assistant — *August 2022 - Present*
    - Conducted data analysis and pre-processing on a large Protein bank dataset containing over 15,000 rows and 5,000 target classes.
    - Implemented and compared multiple machine learning and deep learning algorithms, including Random Forest, XGBoost, and Neural Networks, to predict protein structures and classify proteins with 95% accuracy on the test dataset.
    - Demonstrated strong ML expertise and effective communication by presenting results through interactive PowerBI visualizations to the project supervisor.
    - Proactively developed skills by exploring new algorithms and techniques to drive team success.

- **Chartered Professional Accountants of British Columbia** — Burnaby
  Data Scientist Consultant — *Jan 2022– April 2022*
    - Utilized my Python skills to implement effective data cleaning and pre-processing techniques on a large dataset containing information on over 50,000 members, resulting in improved data quality.
    - Leveraged Power BI and ArcGIS (geographic information system) to create visually impactful visualizations and maps that revealed key target areas for recruitment and demographic insights of members.
    - Applied advanced regression algorithms, including ridge, KNN, lasso, and Linear, to accurately predict the number of members that would join over the next five years.
    - Presented strategic recommendations to the Vice President of CPABC, including a 5-year plan for resource allocation, recruitment, and data storage. My suggestions resulted in a reduction of 50 hours per week and a cost savings of 15-20% in database management.

PROJECTS

- **Telecom Churn Prediction (Classification, EDA, Data cleaning, Data pre-processing, Streamlit, Railway)**:

  Developed a predictive model to analyze customer behavior and predict churn using Python and machine learning techniques. Deployed the model as a Streamlit app and hosted on Railway, achieving an accuracy of 95% on the test dataset.

- **Hate Speech Classification via Tweets (NLP. Spacy, NLTK, Deep learning, Sentiment Analysis, Streamlit, Railway)**:

  Developed an NLP-based hate speech classifier for tweets using Python and various ML algorithms (Naive Bayes, Decision tree, Random Forest, SVM, deep learning), achieving 96% accuracy. Deployed as a Streamlit app and hosted on Railway.

- **Spotify playlist recommendation (Spotipy API, PCA, Dimensionality reduction, K-means clustering)**:

  Developed a music recommendation system using Spotify's API (Spotipy) and Python, utilizing K-Means Clustering. Provided links to the 11 playlists to the end-users. Achieved a high user satisfaction rate by providing personalized and diverse playlists.