SUBASH KHANAL

(977) 98650 03545 | Subashkhanal2580@gmail.com
https://www.linkedin.com/in/subash-khanal-004521211/
https://github.com/Reddevil2580
Kathmandu,Nepal

Skills

- SQL
- Python
- Microsoft Excel
- C/C++
- Web scrapping
- Data Analysis
- Data Visualization
- Statistics
- NLP

Projects

KITWE NEWS AGGREGATOR - OMDENA PROJECT — Present

- Designed and implemented a **data preprocessing pipeline** to clean, transform, and prepare over 10,000 news entries for analysis, ensuring high-quality inputs for modeling.
- Applied **Natural Language Processing (NLP)** techniques, including **Latent Dirichlet Allocation (LDA)** for topic modeling and **TextBlob** for sentiment analysis, to extract insights from textual data.
- Created a final labeled dataset with over 10,000 records categorized as fake or real and maintained the balance between the labels using **SMOTE technique** to handle class imbalance ensuring **data integrity** of downstream machine learning tasks.

UNEMPLOYMENT ANALYSIS FROM 1991-2021 - PERSONAL PROJECT — September 2024

- Utilized **Python** to analyze global unemployment rates across countries and continents from 1991-2021.
- Applied **ARIMA** model to forecast Nepal's unemployment rate through 2030, highlighting future economic challenges.
- Analyzed continental trends, identifying Africa with the highest average unemployment rate (9.59%) and Oceania with the lowest (3.85%).

BIBLICAL CHARACTER NETWORK ANALYSIS - COLLABORATIVE PROJECT — July 2024

- Collaborated with a peer to perform network analysis using **NetworkX** on relationships among characters in the Bible by J. Charles.
- Utilized **spaCy** for efficient name extraction from the text, enhancing the accuracy of character identification.
- Implemented **data analysis** techniques to identify the most influential characters, repeated names, and significant connections.
- Compiled findings into a comprehensive report, offering new perspectives on Biblical narrative structure and character importance.

CERTIFICATIONS

DATA ANALYSIS WITH PYTHON – COGNITIVE CLASS IBM — September 2024

- Learned **exploratory data analysis** techniques including descriptive statistics, ANOVA, correlation analysis, and data visualization using Python libraries such as **pandas** and **matplotlib.**
- Developed proficiency in **data wrangling** skills, including handling missing values, data normalization, binning, and converting categorical variables to quantitative formats.
- Implemented various **machine learning models** including linear regression, polynomial regression, and ridge regression, with a focus on model evaluation, refinement, and selection techniques.

PYTHON FOR DATA SCIENCE – GREAT LEARNING — January 2024

- **Developed proficiency in data manipulation and analysis** using NumPy and Pandas, including handling Series and DataFrame objects for efficient data processing and exploration.
- **Learned data visualization techniques** with Matplotlib, creating various plots such as line, bar, scatter, histogram, box, and violin plots to interpret and present data insights effectively.
- **Applied statistical analysis skills** to generate meaningful insights from data, facilitating informed decision-making through visual representation and exploratory data analysis.

Education

BACHELOR OF SCIENCE IN COMPUTER SCIENCE – Kathmandu University – 28 kilo,Dhulikhel Present

Work Experience

DATA SCIENCE INTERN – CODSOFT Feb 2024-Mar 2024

- **Analyzed sales data** using NumPy and Pandas; created scatter plots and heatmaps, and developed linear regression models to predict sales based on advertising spend.
- **Performed survival analysis** on the Titanic dataset; cleaned data, encoded variables, and trained a logistic regression model, providing interactive predictions and evaluating model accuracy.
- **Explored the Iris dataset** with EDA and visualization; identified feature correlations and redundancy, trained a logistic regression model, and achieved high classification accuracy.
- **Executed data preprocessing and modeling** tasks, including handling missing values, encoding data, and saving predictions, showcasing practical skills in data science methodologies.

FREELANCER - OMDENA Oct 2024-present

- Developed a comprehensive **Kitwe News Aggregator** system to classify news as real or fake using advanced NLP techniques.
- Conducted extensive **Exploratory Data Analysis (EDA),** analyzing linguistic patterns, source credibility, and sentiment for effective feature engineering.
- Implemented machine learning models with **pre-trained transformers (BERT, DistilBERT)** to predict news authenticity, achieving high accuracy and F1 scores.
- Deployed the Kitwe News Aggregator as an interactive web application using **Streamlit Cloud**, enabling real-time fake news classification and enhancing accessibility for end-users.