

End-to-End Wikipedia Page Views Prediction

For my capstone project, I developed a machine learning model to predict Wikipedia page views based on historical traffic data. The goal was to create a model that can forecast future page views, helping users predict how much traffic a given page might receive on a specific day. This type of time-series forecasting is especially useful for websites and platforms that need to forecast traffic for better resource planning, content creation, and trend analysis.

The dataset I used was sourced from Kaggle, containing daily views for thousands of Wikipedia pages over several years. I initially explored models like ARIMA, Prophet, and XGBoost, but ultimately decided to use a deep learning model—specifically, a feedforward neural network (FNN) with Sequential and Dense layers. This approach was simple yet effective for time-series data, and it performed well without overfitting. As someone new to machine learning, I felt most comfortable with this method because it was straightforward to implement and adaptable, giving me solid results without unnecessary complexity.

After training the model, I built a Flask app that allows users to select a Wikipedia page and see the predicted page views for the next day. The app also includes features like determining whether the predicted traffic is high, low, or average compared to the past 30 days, and it provides a confidence score that indicates how certain the model is about the prediction.

I deployed the app on Replit, making it publicly accessible via a link for anyone to interact with. Additionally, I documented the entire process on GitHub, providing all the code, deployment instructions, and a README to guide anyone who wants to run the app locally.

This project showcases how machine learning can be applied to real-world time-series forecasting problems, offering meaningful insights into traffic patterns and predictions. I'm proud of what I've built and excited to continue exploring how to create more practical solutions with machine learning.