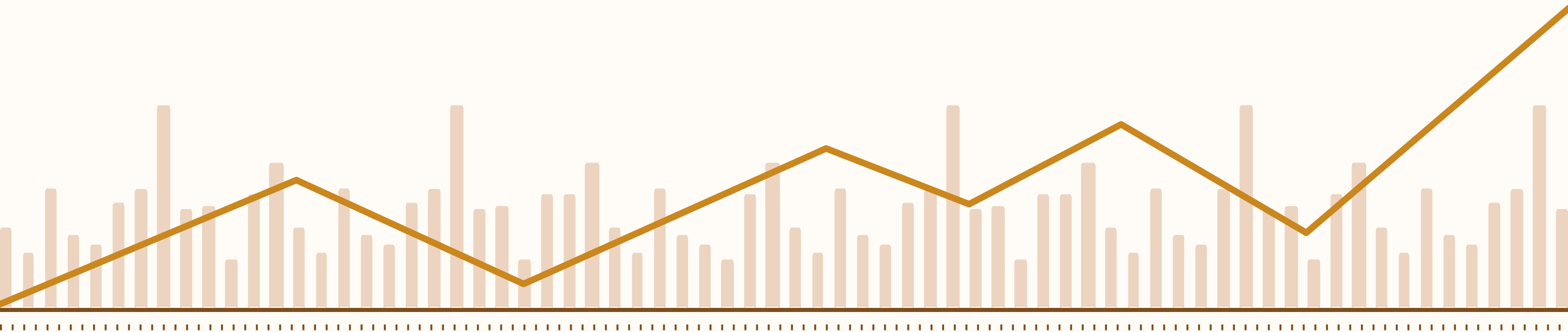


Net Asset Value Forecasting

Providing Tailored insights for Forecasting Net Asset Values Using the power of Machine Learning

Group 4 Team

Asya Haji

Gabriel D

Gloria M

Karim N

Watson

1

Introduction

Introduction

Net Asset Value:
Net asset value is the value of an entity's assets minus the value of its liabilities, often in relation to open-end, mutual funds, hedge funds, and venture capital funds.

Problem Statement:
How can we accurately forecast future NAVs for various schemes using historical data in order to strategically guide the investment company and inform investors about potential returns?

The Main Objective

The main objective is to help Organizations (UTT AMIS) or individuals make informed decisions on investment portfolios

This will be done by creating a robust forecasting model that will leverage historical data to make predictions on the Net Asset Values of a certain scheme after a certain period of time.

By identifying the future Net Asset Values Organizations or individuals can be able to make informed decisions on whether or not to invest on a certain scheme.

Specific Objectives

Data Collection

Obtaining data from the UTT AMIS organization website and put it a form which can be utilized to create the model (s).

EDA

Exploratory Data Analysis is done in order to obtain insights and capture trends that are available in the data.

Visualization

Visualizations are also done in order to obtain clearer insights of the data in forms of plots and graphs.

Feature Engineering

The Task of feature engineering is accompanied with feature selection and feature importance, all in order to identify the best features to be used.

Model Development

The task of model development is done by selecting and continuously testing various models and algorithms for both ML and Timeseries

Model Deployment

The task of model deployment was done using streamlit and hosting using the streamlit cloud community.

The Project Impact

Data-Driven Investment Decisions

The accurate forecasting of NAV prices empowers investment firms to make informed decisions based on data, rathFor the "Project's Impact" slide related to your project on forecasting NAV prices across different investment schemes,

Tailored Investor Recommendations

The project enables firms to provide personalized recommendations to investors based on forecasted market behaviors, aligning with individual investment goals.

Risk Management

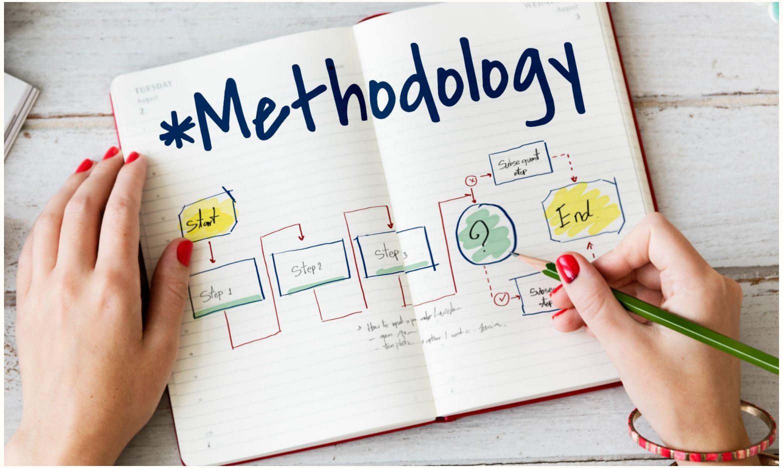
Accurate forecasts can also serve as a risk mitigation tool, allowing firms to anticipate market downturns and adjust their strategies accordingly.

Strategic Marketing

With knowledge of upcoming NAV trends, marketing campaigns can be strategized more effectively, targeting potential investors at the right time.

2

Methodology



SEMMA

Sample

Extract a subset of data from the UTT AMIS official website for initial exploration.

Explore

Analyse the data to identify patterns, relationships, and anomalies

Modify

Transform the data into a format suitable for model training

Model

Train the machine learning model using the processed data

Asses

Evaluate the model's performance using metrics like MAE, RMSE, MSE and R2.

Iteration

This process explains how we used an iterative approach together with SEMMA

Tools Used

Main Tools

- Computer
- Python
- Jupyter Notebook

Feature Engineering

- Python libraries - sklearn, Random forest.

Presentation and report

- Google slides
- Microsoft Office
- Gmail
- Google drive

EDA

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Pandas profiling

Modelling

- Random Forest
- Ensemble
- LSTM

Model Deployment

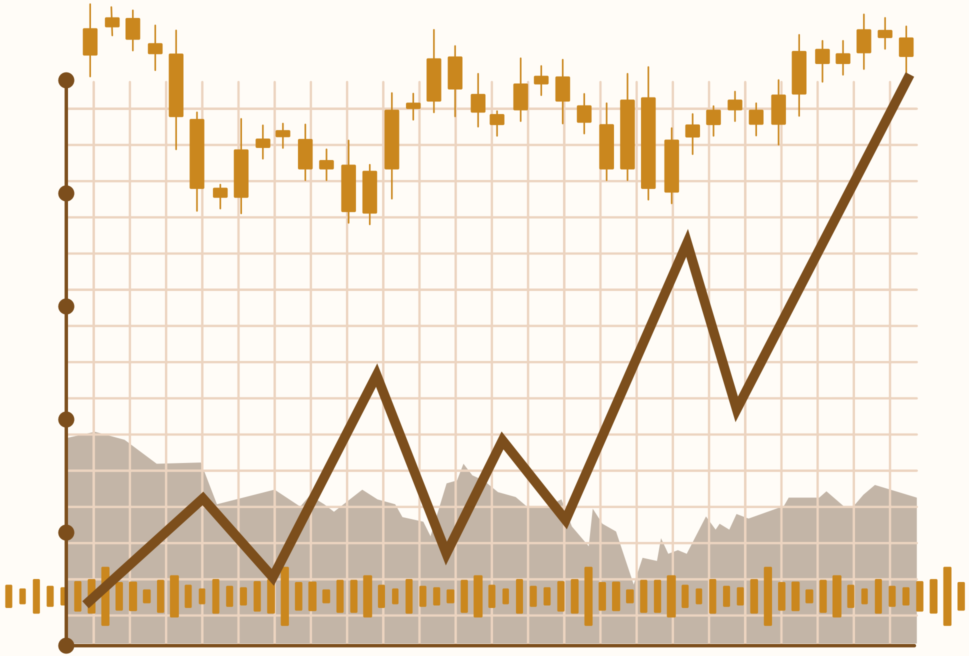
- Streamlit
- Dashplot

Project Implimentation Timeline

- Data Collection and Preparation (Web Scraping).
- Data Preprocessing and Cleaning.
- Exploratory Data Analysis and Data Division.
- Data Visualization.
- Feature Engineering, selection and Importance.
- Model Selection and Model Development.
- Model Evaluation and refining.
- Model Deployment or Hosting.

Data Collection

- **Selenium Automation:** Automated web actions using Selenium framework.
- **Data Extraction:** Directly extracted datasets from UTTAMIS website tables.
- **Dynamic Content Handling:** Overcame challenges posed by dynamic website content.
- **Programmatic Navigation:** Programmed framework for site navigation and data capture.
- **Data Structuring:** Organized extracted data, often as DataFrames.
- **Smooth Extraction:** Ensured seamless data retrieval, including CAPTCHA and load time management.



Data Pre-processing

- **Missing Values:** Identified and filled missing values using mean, median, mode, interpolation, or predictive modeling.
- **Duplicate Entries:** Removed duplicate entries to prevent redundancy.
- **Outliers:** Detected and handled outliers using IQR or Z-scores.
- **Feature Scaling:** Normalized or standardized features for consistent range.
- **Categorical Encoding:** Encoded categorical variables with methods like one-hot encoding.
- **Multicollinearity:** Addressed multicollinearity issues by selecting or transforming features for model accuracy and interpretability.

EDA

- **Key Characteristics Analysis:** Examined dataset distributions and patterns for NAV insights.
- **Visualization:** Used histograms, scatter plots, and correlation matrices to uncover variable relationships.
- **Outlier Detection:** Identified and managed outliers for data integrity.
- **Missing Value Handling:** Addressed missing data to ensure completeness.
- **Bias Assessment:** Evaluated potential dataset biases.
- **Foundation for Forecasting:** EDA informed forecasting model selection by enhancing data understanding.

Data Visualization

- **Trend Analysis:** Used visualizations to identify trends and patterns in NAV data over time, including seasonality and anomalies.
- **Data Exploration and Validation:** Visualizations aided initial data exploration and validated model results through graphical comparisons of forecasts and actual data.
- **Simplified Communication:** Visualizations conveyed complex financial behaviors, making it easier for stakeholders to make informed decisions.

Feature Engineering

- **Feature Enhancement:** Created new features to improve NAV forecasting.
- **Lag Features:** Introduced lag features to capture past NAV patterns.
- **Rolling Statistics:** Utilized rolling statistics to emphasize trends.
- **Interaction Features:** Combined existing data through interaction features.
- **Seasonal and Time-Based Attributes:** Included seasonal and time-based attributes.

Model Selection and Development

- **Problem Type:** Regression (Time Series): The goal is to predict NAV values for the next six months based on historical data.
- **Input Data:** Cleaned and preprocessed historical data is used as input for model training.
- **Model Selection:** The following algorithms were employed for training: Voting Classifier, Random Forest, Lasso Regression and Lasso Ridge.
- **Feature Selection:** During model training, features were selected, depending on the model and scheme, through feature engineering to achieve the best forecasting results.

Model Evaluation

- **Evaluation Metrics:** Various evaluation metrics were employed to assess model performance.
- **Data Splitting:** The dataset was divided into training and testing sets using the "train-test split" method.
- **Best Performing Models:** The top-performing models were identified as follows:
  - LSTM (Long Short-Term Memory).
  - Random Forest.
  - Ensemble model.

Model Evaluation (NAV Per Unit)

	Model	MAE	MSE	RMSE	R-Squared	score
Bond	Ensemble	0.2775	0.1188	0.3446	0.9911	99.11%
Jikimu	Ensemble	3.3717	16.7708	4.0952	0.8674	86.75%
Watoto	Ensemble	17.5572	398.3220	19.9580	0.9374	93.75%
Liquid	Ensemble	5.6915	42.8282	6.5443	0.9858	98.58%
Wekeza Maisha	Ensemble	30.15786	1133.91436	33.6736	0.9405	94.06%
Umoja	Ensemble	23.8508	741.3169	27.2271	0.9401	94.01%

Model Evaluation (NAV)

	Model	MAE	MSE	RMSE	R-Squared	score
Bond	Ensemble	16698.624	376401495.12	19401.069	0.9749	97.49%
Jikimu	Ensemble	3407.682	65437492.9	8089.344	0.36427	36.43%
Watoto	Ensemble	328.217	189031.073	434.777	0.9964	99.64%
Liquid	Ensemble	15127.30	1415785147.2	37626.92	0.9616	96.16%
Wekeza Maisha	Ensemble	2027.2923	6823719.47	2612.225	0.9549	95.50%
Umoja	Ensemble	2742.58	618500901.2	24869.67	0.57316	57.32%

Model Deployment

Deployment involved creating a web application system in which the model(s) can be used to predict the **outcome (Net Asset Value)** based on the given inputs.

The application required online hosting for any user to access via a link. The application was hosted on the streamlit cloud community.

Link

nav-forecasting-uttamis.streamlit.app

Thanks!

Do you have any questions?