**ENCRYPTON HACKATHON**

**Team Solomon:**

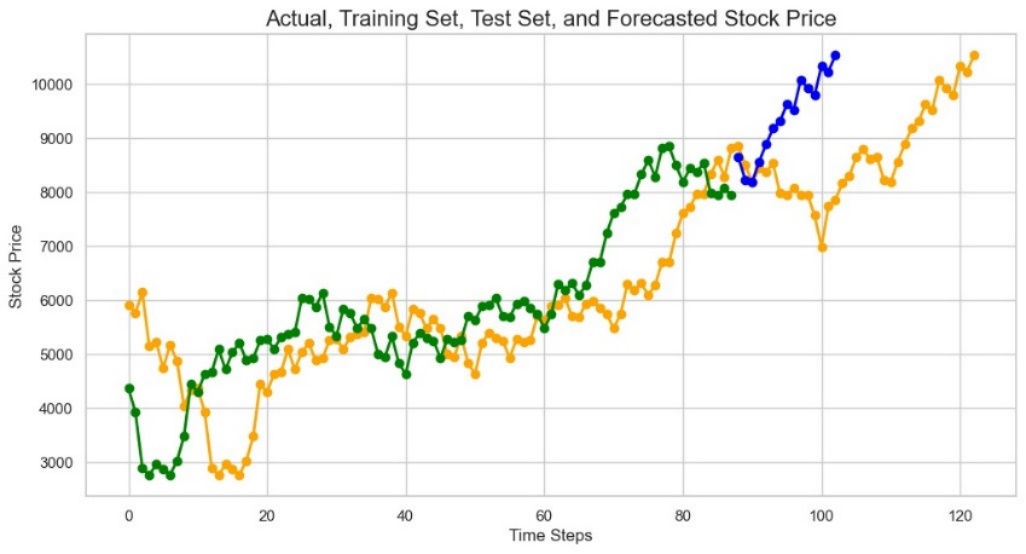
**Taneti Sanjay,**

**Narne Dheeraj Balaram.**

**Problem Statement:**

To create innovative AI and machine learning models that can predict trends and risks in green finance, helping financial institutions make informed and sustainable decisions.

|  |  |
| --- | --- |
| **Abstract**  This study introduces an innovative approach to address the growing importance of sustainability in financial decision-making through the integration of artificial intelligence (AI) and machine learning (ML) models. The proposed solution aims to predict trends and assess risks in the realm of green finance by synergizing financial and environmental data. By leveraging advanced analytics techniques, our methodology provides financial institutions with a comprehensive and forward-looking perspective on sustainable investments.  The integration of Environmental, Social, and Governance (ESG) principles forms the foundation of our approach, ensuring a holistic evaluation of investments. Through the fusion of financial insights and environmental metrics, our models offer a robust framework for predicting future trends and identifying potential risks associated with green finance. This research contributes to the evolving field of sustainable finance by empowering financial institutions to make informed and responsible decisions, aligning their portfolios with long-term environmental and financial goals. The outcomes of this study hold significant implications for the evolutionof green finance, fostering a more sustainable and resilient financial landscape.  **Methodology: Synergizing Financial and Environmental Data**  This section will delve into the technical detail on the integration of financial and environmental data, along with the application of advanced analytics techniques, forms the basis of our innovative AI and machine learning models. It will highlight the specific methodologies used to predict trends and assess risks in the field of green finance, emphasizing the role of Environmental, Social, and Governance (ESG) principles in our holistic evaluation of investments.  **2. Model Building:**  **Simple RNN:**   * 1. **Sequence Creation:**   To prepare the data for training, we transformed the normalized time series data into sequences. Each sequence consisted of 10 consecutive closing prices, and the corresponding target for each sequence was set as the next closing price. This sequence creation is vital for the model to understand the temporal patterns in the data.   * 1. **Model Architecture:**   Our predictive model utilized three recurrent layers: SimpleRNN, LSTM, and GRU. These layers are designed to capture sequential dependencies in the data. Each layer had 50 units, and the Rectified Linear Unit (ReLU) activation function was applied to introduce non-linearity. Reshape layers were incorporated to adjust the output shapes between the recurrent layers. The final layer, a Dense layer with one neuron, represented the predicted price.  **c. Model Compilation:**  Before training the model, we needed to compile it. This involved specifying the optimizer, loss function, and metrics. We chose the Adam optimizer with a learning rate of 0.0001, Mean Squared Error as the | As from the abstract we have divided this problem into two parts one is to predict the trends in green finance and the other is to predict the risks of it.  **Prediction of financial Data:**  In our approach to predicting trends in green finance, we considered all the companies in the NIFTY 50 index. This ensures a comprehensive representation of companies in our study. We used tools such as ‘yfinance’ API to get historical data for these companies (NIFTY 50) inclusive. Following this here are our preprocessing techniques:  **1. Data preprocessing:**   1. **Minmax Scaling:**   We normalized each column present in the data set to be between the values of 0 and 1. As this would aid us a lot when training the model and reduce the complexity of the model.   1. **Principal Component Analysis (PCA) :**   Principal Component Analysis (PCA) plays a crucial role in enhancing the effectiveness of machine learning models by decomposing features and extracting essential information. This process enables a more refined understanding of the inherent patterns within the data, empowering models to make more accurate predictions or forecasts. The feature extraction facilitated by PCA contributes to improved model learning and generalization.  **Note: In our specific implementation, PCA was omitted due to the small dataset size. Using PCA on a smaller dataset could lead to overfitting, where the extracted features may capture noise rather than meaningful patterns, impacting the model's generalization and predictive performance. Caution and thorough evaluation are crucial before applying PCA to datasets of limited scale.**  loss function, and accuracy as a metric for monitoring during training.  d. **Training:**  The model was then trained using the created sequences and target values. Training involved running through the dataset for 1000 epochs with a batch size of 32. This process allowed the model to learn the patterns within the sequences and improve its predictive capabilities over time. We also used a portion of the data for validation during training to monitor the model's performance on unseen data and prevent overfitting.  e. **Prediction:**  After training, we used the model to predict the next closing price based on the last sequence in the normalized data. This involved reshaping the last sequence, inputting it into the model, and obtaining the predicted price. To make the prediction meaningful, we inverted the normalization, bringing the predicted price back to the original scale. |



**-o- Actual Values -o- Testing Values -o- Forecasted Values**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Inferences:**  From our observations, it's evident that the model effectively captures patterns in stock prices, enabling accurate forecasts. The model's proficiency in learning and predicting similar data patterns indicates its capability to anticipate potential price movements. By leveraging these learned patterns, the model demonstrates a commendable performance in familiarizing itself with the intricacies of stock price dynamics and making reliable predictions.  As the dataset which we are using is small (we have to adjust the data to the ESG dataset which we will be discussing in a small while.)  **1. Limited Representativity:** Small datasets may lack diversity, hampering the model's ability to capture a broad range of patterns.  **2. Overfitting Risk:** With insufficient data, there is a higher chance of overfitting, where the model overly tailors itself to the limited training examples.  **3. Model Complexity Concerns**: Complex models may memorize the small dataset, compromising their generalization capabilities.  **4. Uncertain Predictions**: Limited data availability often results in less confident predictions, introducing uncertainty into the model's output.  **To overcome these problems, we can use the following methods:**  1. Augmentation of datasets.  2. Implement techniques such as dropout or L1/L2 regularization.  3. Choosing a simpler model rather than a complex one.  **The outcomes of the predictions are as follows:**  **1. Simple RNN**  While feeding the data to an RNN we have converted the data into sequence data.    **2. Linear Regression**    **3. Random Forest**    improve model robustness and accuracy in resource-constrained situations. The choice depends on the task's characteristics and requirements. | **Prediction of Risk factor**  We build multiple models to evaluate and forecast Risk Factor.  We will be going over step by step on how the prediction or forecasting is done for the risk factor.  **1. Data Preparation:**  **NOTE: HIGHER THE RISK FACTOR THE BETTER THE INVESTMENT IS (SORRY FOR THE CONFUSION).**  In our risk assessment framework for investment decisions, we utilized an Environmental, Social, and Governance (ESG) dataset specific to India, meticulously sourced from the World Data Bank. This dataset encompasses diverse factors contributing to ESG considerations. Employing a judiciously chosen mathematical function, specifically the mean, we derived an ESG rating or risk factor.  This resulting risk factor serves as a pivotal metric, offering nuanced insights into the sustainability profile of investments. A lower risk factor indicates non-sustainable investments, providing crucial discernment for investors. The predictability inherent in the risk factor underscores that a lower value signifies higher potential risks, prompting a judicious approach to investment decisions.  Our methodology integrates quantitative analysis with ESG principles and introduces a predictive dimension. This elevates our risk assessment approach, enabling stakeholders to proactively navigate potential risks associated with non-sustainable investments. Consequently, our method empowers investors with a robust framework for making informed decisions in the dynamic landscape of financial considerations and sustainability imperatives.  The mathematical function which we chose to use was mean.  **4. ARIMA**    **Inferences:**  The analysis reveals that Random Forest and Linear Regression effectively capture prevalent data patterns. RNN exhibits potential with increased data and training time. Our ranking for the forecasting task, based on collective evaluation, is as follows: Random Forest and Linear Regression performed notably, while RNN shows promise with extended data and training duration.   |  |  | | --- | --- | | 1 | Random Forest | | 2 | Linear Regression | | 3 | RNN | | 4 | ARIMA |   **Cons of our approach:**  The dataset we used consisted only data from 2000 to 2017 that too yearly data. We tried so hard for a dataset which would fit our solution, but we failed to do so. But still we decided to use some mathematical formulae to get monthly data and continue with the solution. Our suggestion is to get a good and trustworthy data set and work on it. The results would still be similar to this.  **Overcoming the problems:**    To address issues arising from insufficient labelled data or domain-specific challenges, strategies like transfer learning and utilizing pre-trained Generative Adversarial Networks (GANs) can be employed. Transfer learning involves adapting a model pretrained on a related task, leveraging learned features for a new task with limited data. Alternatively, pre-trained GANs, skilled at generating realistic data, can be fine-tuned, or used as feature extractors for specific tasks, enhancing model performance in scenarios with scarce labelled data. These methods capitalize on existing knowledge to |

**Conclusion:**

Using the predictions made by both of the parts the financial investor will be able to make sustainable investments.

**Future Scope:**

The current implementation represents a prototype rather than a fully completed project. There is significant potential for enhancement and expansion. By establishing a robust data pipeline, the models can undergo continuous training, enabling them to iteratively learn and enhance their predictive and forecasting capabilities over time. This ongoing improvement process opens the door to refining and expanding the project's scope in the future.

**References:**

1. [Short-term stock market price trend prediction using a comprehensive deep learning system](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00333-6) – Jingyi Shen.

1. Top of Form