Enhanced Data-Driven Due Diligence for Unveiling Business Opportunities

S.V.R Vara Prasad [1], Surarapu Vagdevi[2], Yalaka Sai Sri Chandana[3], Dwarampudi Sai Ravi Teja Reddy[4] , Uppalapati Mounika Bindu[5] , Chintala Deepika[6]

[1]Assistant Professor , [2-6] B.Tech Student, LIET
[1,2,3,4,5,6] Computer Science and Information Technology , Lendi Institute of Engineering and Technology, Vizianagaram

--***--

ABSTRACT
This project, titled "Data-Driven Due Diligence - Unveiling Business Opportunities," is dedicated to empowering Indian entrepreneurs in making informed financial decisions. In contrast to existing platforms with prohibitively high subscription costs, our project presents a cost-effective solution utilizing Python technologies, including Streamlit, SQLite, and Machine Learning libraries for linear regression. The web application comprises a streamlined signup/login process, an exclusive listing page for Indian entrepreneurs to input company details, and a versatile investment page accessible to both Indian and foreign entrepreneurs. Notably, SQL is employed for efficient database management. Through the strategic integration of advanced technologies and data analysis, this project equips entrepreneurs with valuable insights to optimize their investment strategies, fostering greater profitability.

*Key Words***:** Data analytics, Due Diligence, Data visualization, Python, Machine Learning, SQLite.

1.INTRODUCTION
In response to the financial constraints faced by Indian entrepreneurs due to high subscription costs on existing platforms, our project, "Data-Driven Due Diligence - Unveiling Business Opportunities," endeavors to revolutionize the landscape of financial decision-making. Acknowledging the pivotal role of entrepreneurs in fostering economic growth, our initiative is tailored to address the unique challenges encountered by Indian entrepreneurs.

Entrepreneurs serve as the driving force behind a nation's economic progress, bringing forth innovative ideas, creating employment opportunities, and contributing substantially to wealth generation [1]. However, the financial barriers hindering their growth cannot be overlooked. Small and medium-sized enterprises (SMEs) play a crucial role in India's economic fabric, and our project is strategically designed to empower these entities.

The financial challenges faced by Indian entrepreneurs are further exacerbated by the prohibitive subscription costs associated with existing platforms. To counter this, our project leverages cutting-edge Python technologies, including Streamlit and machine learning libraries for linear regression, to offer a cost-effective and accessible solution. This initiative aims to democratize data-driven due diligence, ensuring that entrepreneurs, irrespective of their financial capacities, can make informed decisions.

The significance of our project extends beyond mere technological innovation. It aligns with the broader narrative of entrepreneurship in India, recognizing the substantial contributions of small businesses to the nation's economic growth. By providing a user-friendly platform with tailored features, we aim to not only address the financial challenges faced by entrepreneurs but also foster a conducive environment for their sustained success.

As we delve into the project details, technologies used, and results, it becomes evident that our approach goes beyond conventional due diligence. The integration of digital entrepreneurship components further enhances the decision-making capabilities of our target audience, positioning our project as a comprehensive solution for the dynamic landscape of Indian entrepreneurship.

2. Proposed Method
Our project, "Data-Driven Due Diligence - Unveiling Business Opportunities," is meticulously crafted to address the specific needs of Indian entrepreneurs who face financial constraints associated with exorbitant subscription costs on existing platforms. Unlike prevalent websites charging as much as $999 per month, our project is committed to democratizing access to data-driven due diligence for Indian entrepreneurs. Leveraging Streamlit and various Python libraries, including SQLite and a machine learning library for Linear Regression, our methodology aims to provide a cost-effective and user-friendly solution.

2.1 GRAPHICAL ANALYSIS AND PREDICTIONS
A standout feature of our proposed method is the integration of Graphical Analysis and Predictions within the Investment Page. This innovative approach goes beyond traditional data-driven due diligence, providing

users with visually intuitive insights into stock values and future trends.

2.1.1 Graphical Analysis

Upon selecting a company for investment or analysis, users can delve into comprehensive graphical representations of stock values. Interactive charts, including line graphs and candlestick charts, dynamically illustrate the historical performance of the selected company. Users can customize timeframes, compare multiple companies, and gain a deeper understanding of market trends.

2.1.2 Predictions

In addition to graphical analysis, our platform offers a unique three-year rough prediction feature. Powered by machine learning algorithms, this functionality forecasts potential future stock values based on historical data and market trends. While these predictions serve as valuable insights, users are encouraged to interpret them cautiously, recognizing the dynamic nature of financial markets.

This combination of Graphical Analysis and Rough Predictions enhances the decision-making capabilities of entrepreneurs, providing them with a visual and forward-looking perspective on their investments. As an integral part of our project, this feature sets a new standard for user-friendly and insightful data-driven due diligence in the realm of Indian entrepreneurship.

2.2 INTEGRATION OF DIGITAL ENTREPRENEURSHIP

Our project recognizes the vital role of digital entrepreneurship skills. Entrepreneurs need skills that support the identification and exploitation of opportunities based on digital literacy [2]. The project leverages digital technology, including social media platforms, mobile platforms, crowd-funding platforms, artificial intelligence, big data, and cloud, in alignment with the findings of Cavallo et al. (2019) [3]. The use of digital platforms is acknowledged as a powerful tool influencing the growth of new digital-based businesses by creating the ability to scale the business rapidly [4].

2.3 TECHNOLOGIES USED

Our project employs a set of cutting-edge technologies, combining the versatility of Python programming language with advanced libraries and frameworks to create a comprehensive and effective web application.

2.3.1 Python Programming Language

Python serves as the foundational programming language for our project. Known for its readability, simplicity, and extensive community support, Python enables seamless integration with various libraries and frameworks. Its versatility makes it an ideal choice for developing diverse applications, from data analysis to web development.

By leveraging the capabilities of Python, our project not only harnesses the power of a versatile programming language but also taps into the strengths of advanced features that contribute to code efficiency, maintainability, and overall project success. Several languages can be used to automate data analysis but Python is one of the popular languages for doing the automation and the reason being its simplicity [5], flexibility, and vast library support.

2.4 ALGORITHMS USED

Our project integrates advanced algorithms to enhance data analysis and decision-making processes, with a primary focus on providing valuable insights to users. The key algorithm employed in our project is:
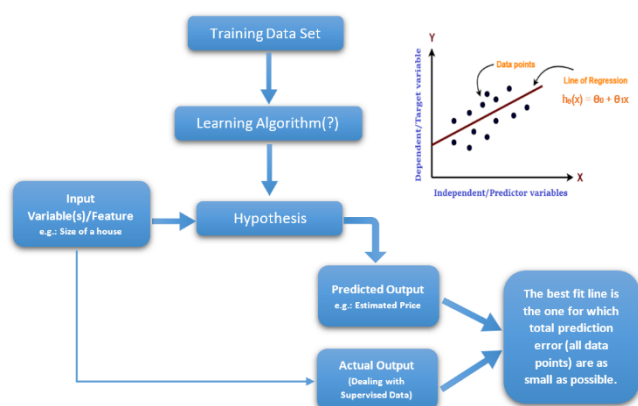
2.4.1 Linear Regression

Linear Regression is a fundamental machine learning algorithm used for predicting a continuous outcome variable based on one or more predictor variables. In our project, Linear Regression is applied to analyze and predict various financial parameters, providing insights into potential investment opportunities.

Linear regression [6] is a typical mathematical research tool that allows you to quantify and simulate expected effects using numerous input variables. The algorithm establishes a linear relationship between the input features and the target variable, allowing us to make predictions based on historical data. In the context of our web application, Linear Regression aids in forecasting stock values and facilitating data-driven due diligence for Indian entrepreneurs.

The steps involved in Linear Regression include:

1. Data Collection: Gathering historical data on stock values and related financial parameters.
2. Data Preprocessing: Cleaning and organizing the data to ensure its suitability for analysis.
3. Feature Selection: Identifying relevant features that contribute to the predictive model.
4. Model Training: Training the Linear Regression model on historical data to establish relationships.
5. Prediction: Utilizing the trained model to make predictions on future stock values.

Fig 2.4.1 *Machine Learning Algorithm Detailed View*

By incorporating Linear Regression, our project provides users with valuable insights into the potential performance of listed companies, supporting informed investment decisions. This algorithmic approach enhances the overall functionality and effectiveness of our data-driven due diligence web application.

2.5 LIBRARIES USED

Our project leverages several powerful Python libraries to create a robust and feature-rich web application tailored for Indian entrepreneurs. Each library plays a specific role in enhancing the functionality and user experience.

2.5.1 Matplotlib.pyplot:

Matplotlib is a comprehensive data visualization library in Python. The pyplot module within Matplotlib provides an interface for creating a variety of plots, charts, and graphs. In our project, Matplotlib.pyplot is employed to generate visually informative graphics, aiding users in the analysis of stock values and market trends.

2.5.2 Matplotlib.ticker:

Matplotlib.ticker is a submodule of Matplotlib that facilitates the customization of tick locators and formatters for axes. This library is instrumental in enhancing the clarity of financial plots by allowing precise control over tick placement and formatting on axes.

2.5.3 SQLite3:

SQLite3 is a lightweight, serverless, and self-contained relational database management system. In our project, SQLite3 is utilized for efficient database management, storing user signup details, investment records, and other essential information. This library ensures the seamless operation of our web application, providing a secure and organized structure for data storage and retrieval.

2.5.4 Streamlit

Streamlit serves as the backbone of our web application, providing a simple and efficient platform for building interactive and data-driven interfaces. Its intuitive syntax and real-time updates make it an ideal choice for creating a seamless user experience.

2.5.5 Pandas

Pandas is a versatile data manipulation library that facilitates the organization, analysis, and manipulation of structured data. In our project, Pandas is instrumental in handling and processing the dataset, ensuring smooth interactions between the user interface and the underlying data. To perform real-world data analysis, pandas [7], a python library is being developed since 2008

2.5.6 Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. In our project, Matplotlib contributes to the graphical analysis feature as shown in **Fig 2.5.6**, enabling the generation of insightful charts and plots.
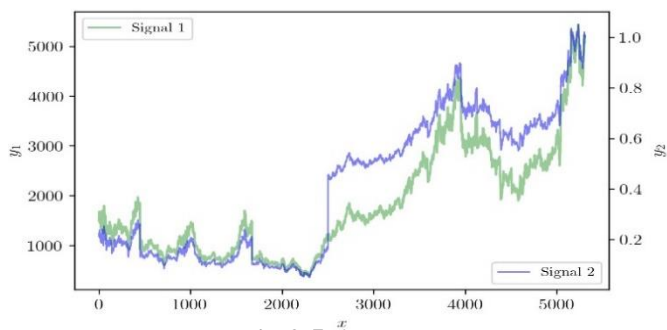
Fig 2.5.6

2.5.7 Scikit-Learn

Scikit-Learn, a machine learning library, provides the Linear Regression model. This model is employed for predictive analysis, generating rough predictions of future stock values based on historical data as shown in the **Fig 2.5.7**. It adds a valuable dimension to the due diligence process. Scikit-learn [8], a Machine Learning library was developed as a Google Summer of Code project.
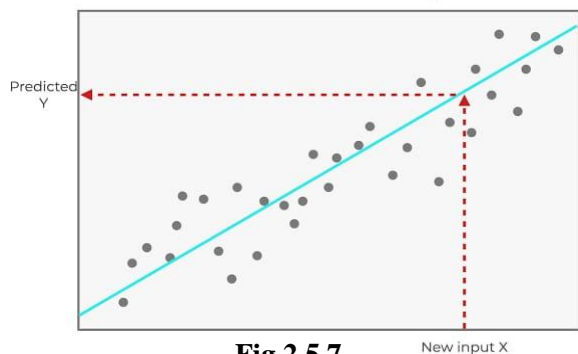
Fig 2.5.7

2.5.8 Base64

Base64 is used for encoding and decoding binary data, and in our project, it plays a role in efficiently handling and presenting certain data formats. This library supports various functionalities, enhancing the overall versatility of our web application.

By harnessing the capabilities of these libraries, our project delivers a dynamic, user-friendly, and data-driven due diligence platform for Indian entrepreneurs.

2.6 RESULTS

The implementation of our project provides a cost-effective and user-friendly solution for data-driven due diligence. The web application facilitates secure user signup, efficient company listing, and seamless

investment processes. Data Analytics mainly deals with the development of computational methods to get insights of data and get intelligent information especially through visualization tools[9].The integration of machine learning for analytics and predictions enhances the decision-making capabilities of entrepreneurs.
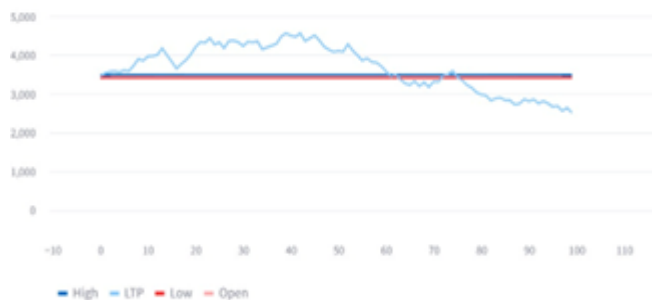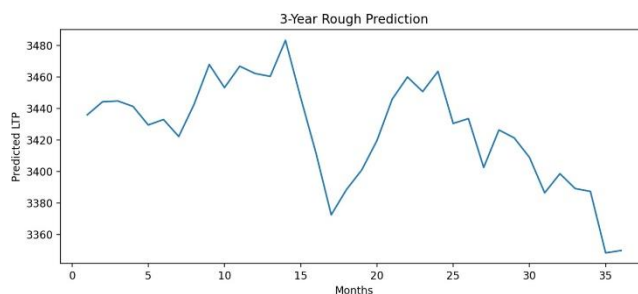

Fig 2.6.1 *Price Movement*
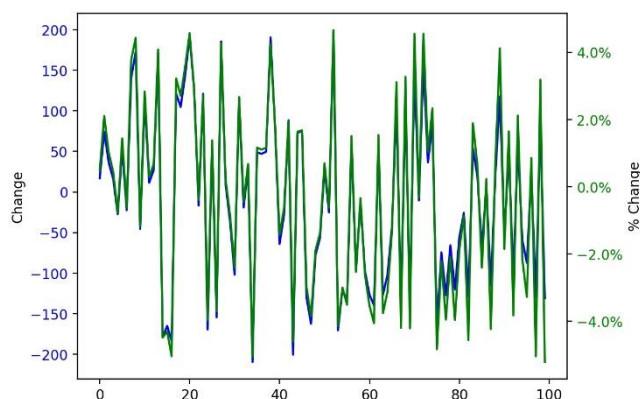

Fig 2.6.2 *Prediction*

Fig 2.6.3 *Daily Change and Percentage Change*

3. CONCLUSIONS

Digital literacy is highlighted as a crucial support for entrepreneurs in discovering and developing relationships and ideas for their businesses [10]. Entrepreneurs often start their businesses driven by passion showcased on digital platforms, with financial motivation becoming a significant factor [10]. The utilization of digital platforms is perceived as a driving force for businesses.

Our project, "Data-Driven Due Diligence - Unveiling Business Opportunities," not only addresses the financial challenges faced by entrepreneurs but also aligns with the significant contribution of entrepreneurship to India's economic progress. By providing a user-friendly platform with tailored features, we aim to empower entrepreneurs to make informed investment decisions. Future iterations will focus on user feedback and continuous improvement.

ACKNOWLEDGEMENT

REFERENCES

1. M. U. Ahmed, "Unleashing entrepreneurship growth in Bangladesh: Why and how", TBS News.
2. T. Bialeti, "Technology Entrepreneurship: Overview Definition and Distinctive Aspects", Technology Innovation Management Review, vol. February, pp. 5-12, 2012.
3. A. Cavallo, A. Ghezzi, C. DellEra and E. Pellizzoni, "Fostering digital entrepreneurship from startup to scaleup: The role of venture capital funds and angel groups", Technological Forecasting and Social Change, vol. 145, pp. 24-35, Aug. 2019.
4. D. M. Steininger, "Linking information systems and entrepreneurship: A review and agenda for IT-associated and digital entrepreneurship research", Information Systems Journal, vol. 29, no. 2, pp. 363-407, Mar. 2019.
5. P. Zhang, Y. Gao and X. Shi, "QuantCloud: A Software with Automated Parallel Python for Quantitative Finance Applications", 2018 IEEE International Conference on Software Quality Reliability and Security (QRS). IEEE, Jul. 2018.
6. Y. Jiang, Y. He and H. Zhang, "Variable selection with prior information for generalized linear models via the prior LASSO method", Journal of the American Statistical Association, vol. 111, no. 513, pp. 355-376, 2016.
7. W. McKinney, "Pandas: a foundational Python library for data analysis and statistics", Python for High Performance and Scientific Computing, pp. 1-9, 2011.
8. F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: Machine Learning in Python", Journal of Machine Learning, vol. 12, pp. 2825-2830, 2011.
9. X. Cai, H. Langtangen and H. Moe, "On the Performance of the Python Programming Language for Serial and Parallel Scientific Computations", Scientific Programming, vol. 13, no. 1, pp. 31-56, 2005.
10. A. S. T. Olanrewaju, N. Whiteside, M. A. Hossain and P. Mercieca, "The influence of social media on entrepreneur motivation and marketing strategies in a developing country", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11195, pp. 355-364, 2018