Application of Machine learning model on stock price prediction

Internship report submitted in fulfillment of the requirements for the Degree of

BTech
in
Computer Science Engineering

BY

MUDIT GARG
20103109

Department of Computer Science Engineering and Information Technology
JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY
(Declared Deemed to be University U/S 3 of UGC Act)
A-10, SECTOR-62, NOIDA, INDIA

AUG - 2023

ACKNOWLEDGEMENT

I would like to place on record my deep sense of gratitude to **Dr. Janardan Verma**, **Assistant Professor**, Jaypee Institute of Information Technology, India for her generous guidance, help and useful suggestions.

I express my sincere gratitude to **Dr. Vikas Saxena (HOD), Dept. of CSE/IT** Jaypee Institute of Information Technology, India, for his stimulating guidance, continuous encouragement, and supervision throughout the course of present work.

MUDIT GARG (20103109)

-

TABLE OF CONTENTS

PROJECT OBJECTIVE

The relentless evolution of financial markets has spurred the exploration of innovative methods to predict stock prices, enabling informed decision-making for investors and traders. In this context, this abstract presents a concise overview of the application of Long Short-Term Memory (LSTM) neural networks in the realm of stock price prediction.

The primary objective of this project is to design and evaluate a robust and accurate stock price prediction model by leveraging the power of Long Short-Term Memory (LSTM) neural networks. This project aims to harness the potential of advanced machine learning techniques to enhance the accuracy and reliability of short-term stock price predictions, contributing to more informed decision-making for investors and traders in the financial markets. Rigorous model evaluation techniques are discussed, shedding light on performance metrics such as Root Mean Squared Error (RMSE), and AIC.

In the ever-changing landscape of financial markets, LSTM-based stock price prediction models stand as a testament to the symbiotic relationship between advanced machine learning techniques and the quest for more accurate, actionable insights into stock price behavior.

PROJECT DESCRIPTION

The "Application of Machine learning model on stock price prediction" project aims to develop an advanced machine learning model that utilizes Long Short-Term Memory (LSTM) neural networks to predict the closing prices of stocks. Closing prices are a crucial indicator in financial markets, influencing investment decisions and strategies. This project seeks to harness the power of sequential data analysis to enhance the accuracy and reliability of short-term stock price predictions, catering to the needs of investors, traders, and financial analysts.

- **Model Developmen**t: Design and implement an LSTM-based machine learning model capable of learning and predicting stock closing prices based on historical data.

- **Data Preparation**: Gather comprehensive historical stock price data, including features such as open prices, high prices, low prices, trading volume, and relevant market indicators. Preprocess the data to ensure its quality, completeness, and suitability for training the LSTM model.

- **Feature Engineering**: Engineer meaningful features from the raw data to capture relevant patterns and trends. Incorporate technical indicators, such as moving averages and relative strength index (RSI), along with sentiment analysis data if available.

- **Training and Validation**: Train the LSTM model on historical data, employing a time-series split validation strategy to assess its performance. Evaluate the model's accuracy using metrics like Root Mean Squared Error (RMSE).

- **Interpretability and Insights**: Analyze the LSTM model's predictions to gain insights into the driving factors behind the stock price movements it predicts. Identify scenarios where the model excels and where its predictions may be less reliable.

CONTRIBUTION

During the period spanning from June 7, 2023, to July 31, 2023, my engagement centered around the captivating realm of stock market closing price prediction through the lens of machine learning techniques. This journey commenced on June 7 with the discovery of an article titled "Stock Closing Price Prediction using Machine Learning Techniques." The article, found at a notable resource, Analytics Vidhya, ignited my exploration.

I learned from a step-by-step implementation guide from Analytics Vidhya, an informative resource elucidating machine learning's application in stock market prediction.

The subsequent days saw a systematic progression of my understanding. I explored diverse regression techniques that underpin prediction methodologies. This knowledge was complemented by a comprehensive study of time series analysis specifically tailored to the dynamic landscape of the stock market.

I explored the intricacies of AutoRegressive (AR), Moving Average (MA), AutoRegressive Moving Average (ARMA), and AutoRegressive Integrated Moving Average (ARIMA) models. This exploration was fortified by a resourceful article, "Stock Market Forecasting Using Time Series Analysis with ARIMA Model," sourced from Analytics Vidhya.

I advanced into the practical implementation of the ARIMA model. To ground my studies, I anchored my analysis in the various datasets.

I learned the intricacies of the Long Short-Term Memory (LSTM) algorithm. This deep understanding was cultivated with insights from a comprehensive article by Intellipaat.

This journey into the intersection of machine learning and finance culminated in the focused study of the application of LSTM in stock market prediction, documented in a scholarly article from ScienceDirect.

Continuing this analytical journey, I did a meticulous comparison between the ARIMA model and LSTM, benchmarked on the basis of the Root Mean Squared Error (RMSE).

The exploration of data sources took center stage on July 1, as I familiarized myself with Yahoo Finance for datasets, followed by an immersion into dataset retrieval and cleaning procedures.

From July 2 to July 12, my focus shifted to real-world datasets from influential companies: ADOBE, AMAZON, UBER, and WALMART. I meticulously worked with essential columns like

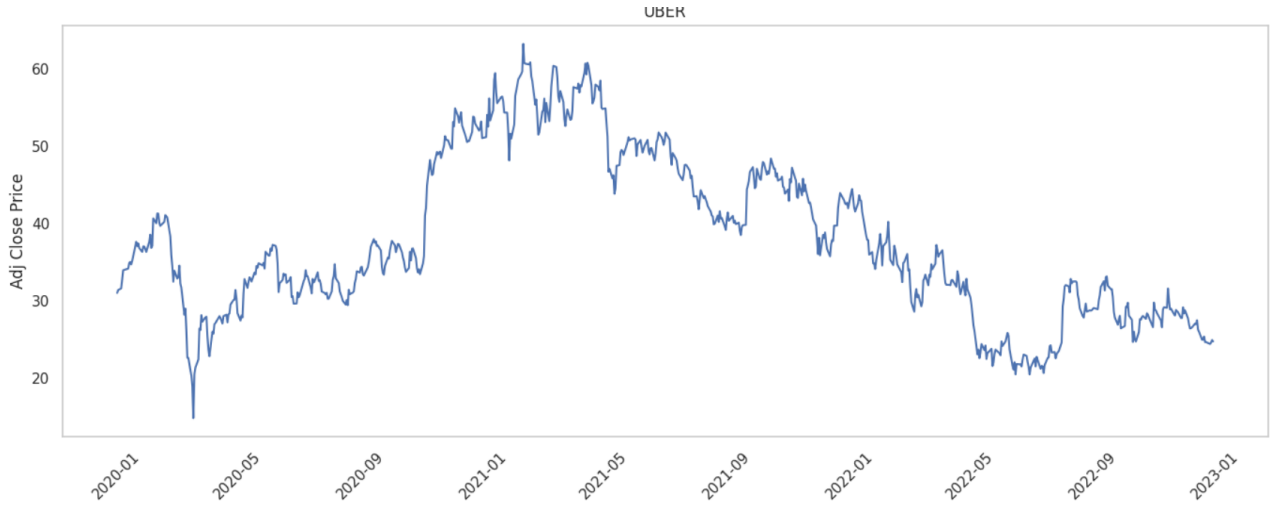
DATE, OPEN, HIGH, LOW, ADJ CLOSE, and VOLUME, investigating differences between OPEN and ADJ CLOSE prices and the nuances of stock volumes.

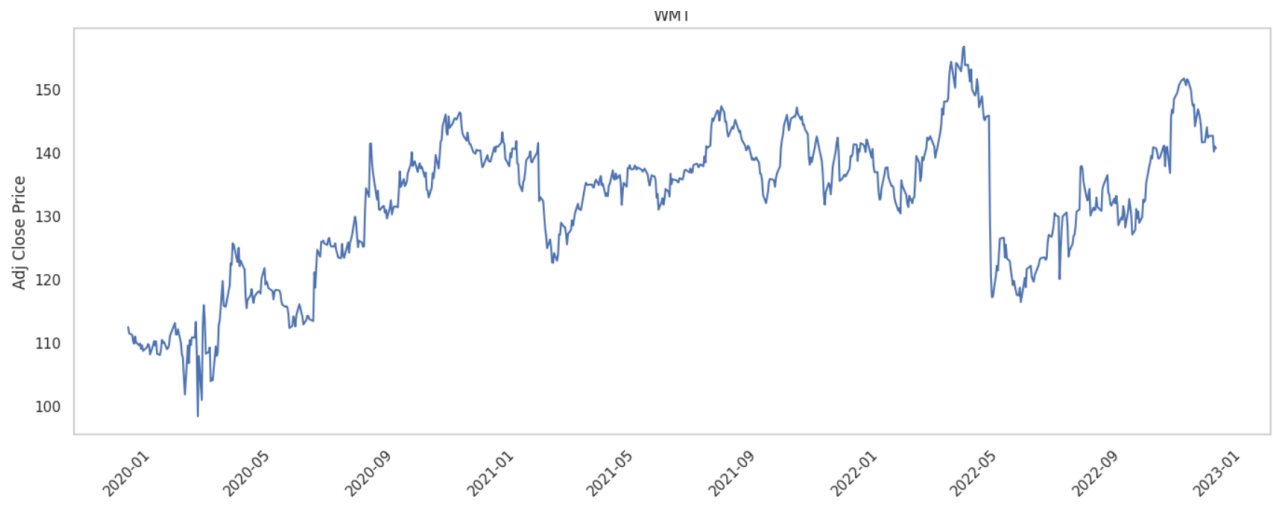
Addressing coding challenges, I resolved keyword errors, name errors, and directory issues, ensuring a smooth workflow.

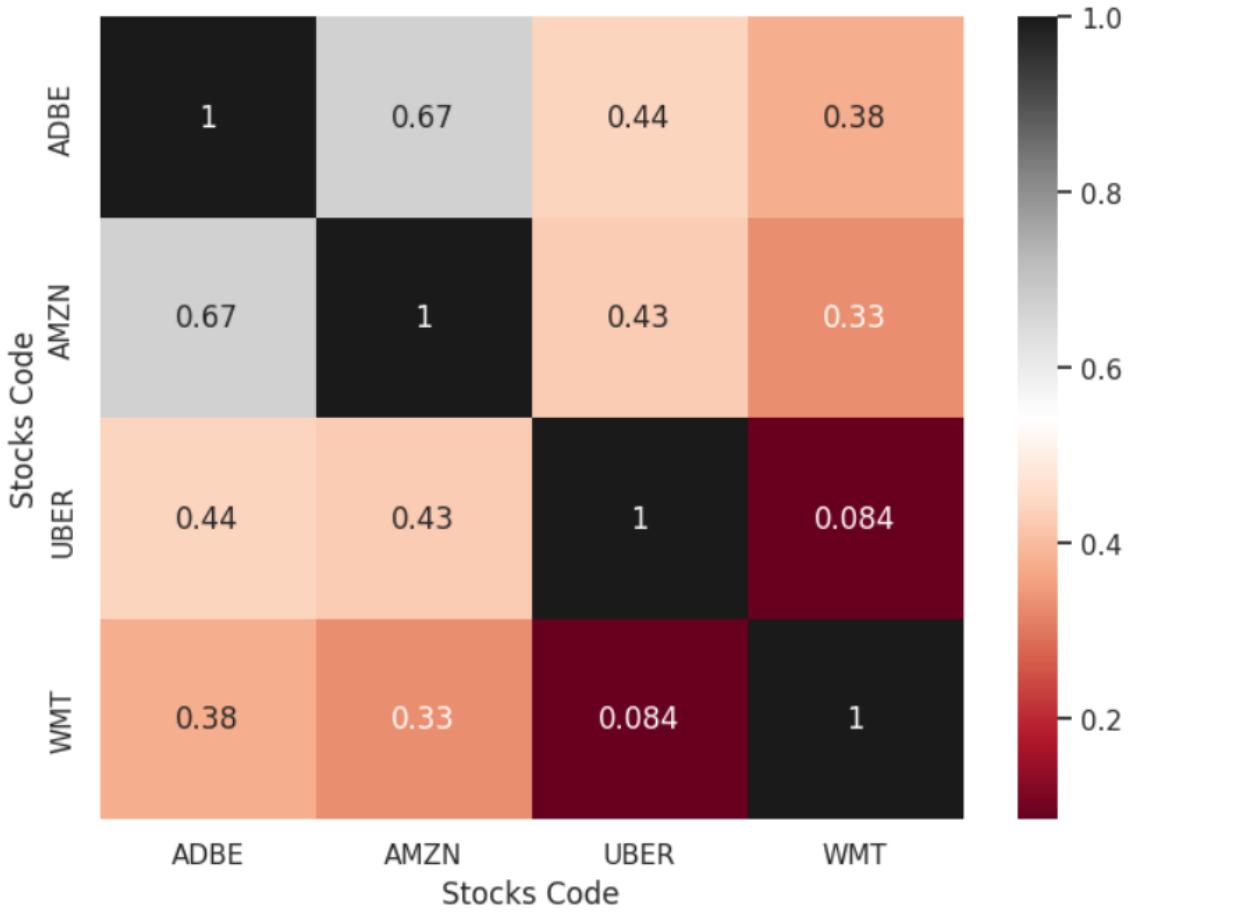
From July 12 to July 22, I delved into the intricacies of data analytics. This phase encompassed calculating and understanding the monthly average of closing prices and volumes, discerning correlations within stock prices, performing moving average calculations, employing Isolation Forest models for outlier detection, constructing training and testing sets for prediction, training LSTM models, and ultimately forecasting values.

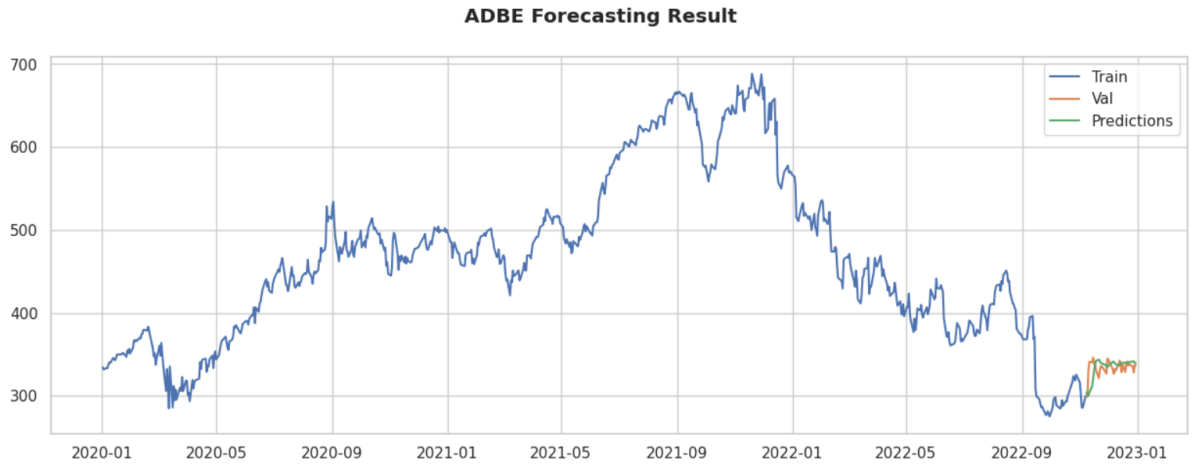
In summary, during this internship I learned the intricate interplay between machine learning and the complex world of stock market prediction, marking a substantial step toward informed financial decision-making.

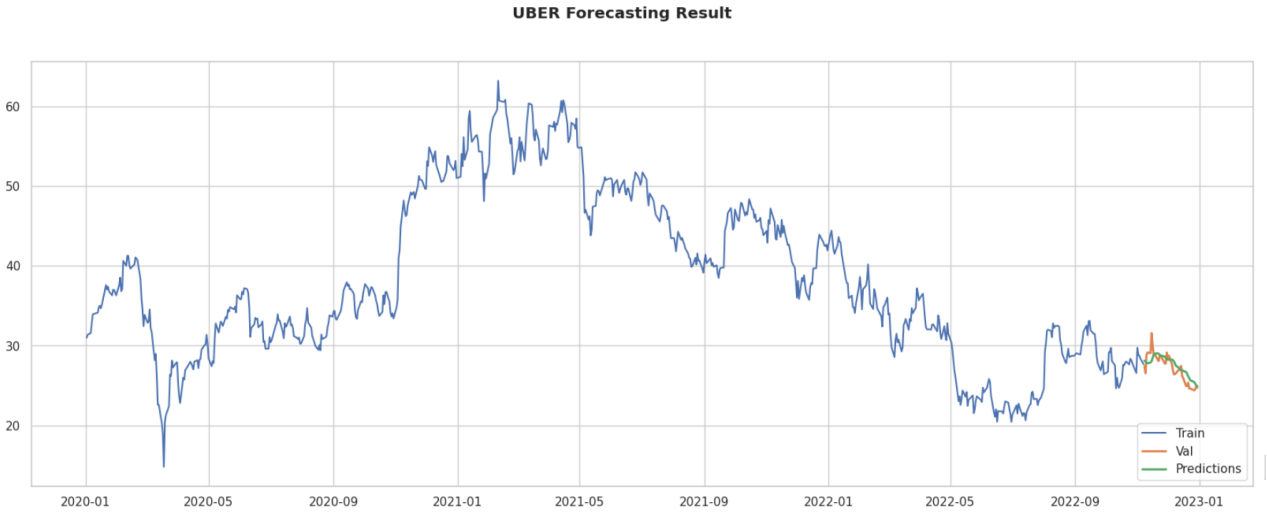

AMZN
Adj Close Price
180
160
140
120
100
80
2020-01
2020-05
2020-09
2021-01
2021-05
2021-09
2022-01
2022-05
2022-09
2023-01

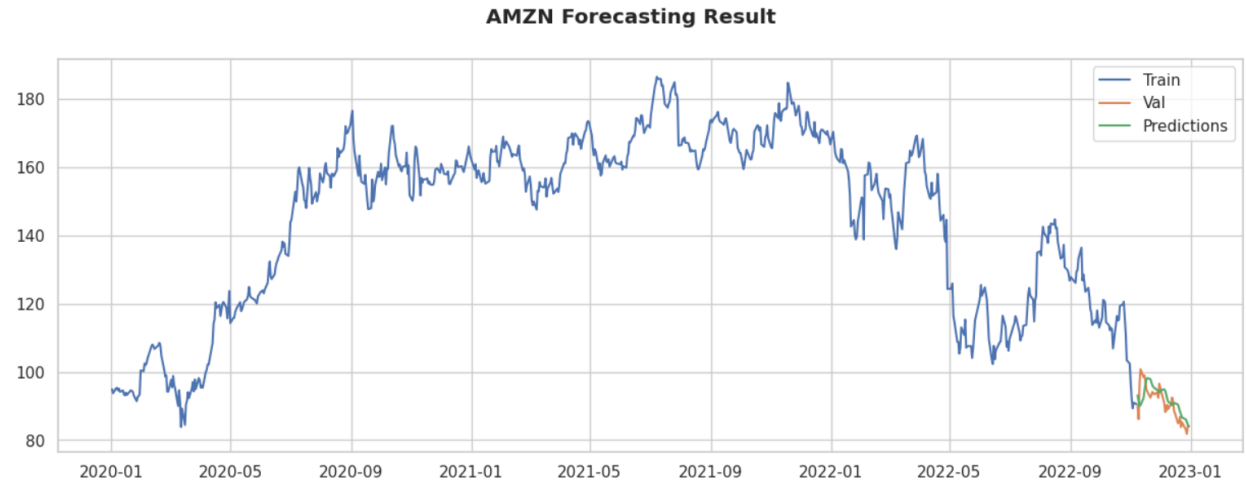

UBER
Adj Close Price
60
50
40
30
20
2020-01
2020-05
2020-09
2021-01
2021-05
2021-09
2022-01
2022-05
2022-09
2023-01


WMT
Adj Close Price
150
140
130
120
110
100
2020-01
2020-05
2020-09
2021-01
2021-05
2021-09
2022-01
2022-05
2022-09
2023-01


Stocks Correlation Based on Percentage Change of Closing Price
ADBE
AMZN
UBER
WMT
Stocks Code
1
0.67
0.44
0.38
0.67
1
0.43
0.33
0.44
0.43
1
0.084
0.38
0.33
0.084
1
ADBE
AMZN
UBER
WMT
Stocks Code
1.0
0.8
0.6
0.4
0.2


ADBE Forecasting Result
Train
Val
Predictions
700
600
500
400
300
2020-01
2020-05
2020-09
2021-01
2021-05
2021-09
2022-01
2022-05
2022-09
2023-01


UBER Forecasting Result
60
50
40
30
20
Train
Val
Predictions
2020-01
2020-05
2020-09
2021-01
2021-05
2021-09
2022-01
2022-05
2022-09
2023-01


AMZN Forecasting Result
Train
Val
Predictions
180
160
140
120
100
80
2020-01
2020-05
2020-09
2021-01
2021-05
2021-09
2022-01
2022-05
2022-09
2023-01

SKILLS DEVELOPED

- **Machine Learning Fundamentals**: Through this project, I gained a deeper understanding of fundamental machine learning concepts, including supervised learning, regression analysis, and sequence prediction, which are essential for creating predictive models.

- **Time Series Analysis**: I developed expertise in time series analysis techniques, enabling me to explore historical stock price data, identify trends, seasonality, and temporal patterns, and incorporate them into the LSTM model.

- **Long Short-Term Memory (LSTM)**: I acquired proficiency in LSTM networks, a specialized type of recurrent neural network (RNN) that excels at capturing long-range dependencies in sequential data. This knowledge allowed me to leverage LSTM's capabilities for stock price prediction.

- **Data Preprocessing**: I mastered data preprocessing techniques, including handling missing values, normalizing data, and ensuring the data is in the appropriate format for training the LSTM model.

- **Real-Time Data Integration**: I learned how to integrate real-time stock price data feeds into the prediction model, enabling the model to provide up-to-date and accurate predictions in dynamic market conditions.

- **Code Debugging and Troubleshooting**: Addressing various coding challenges, including keyword errors, name errors, and directory issues, enhanced my ability to debug code and troubleshoot technical problems.

- **Data Visualization**: I acquired skills in creating informative visualizations to better understand patterns in stock price data, which aided in data exploration and model interpretation.

- **Research and Learning**: Throughout the project, I honed my research skills, finding and

synthesizing information from various sources, including articles, tutorials, research papers, and videos, to enhance my understanding of LSTM and stock price prediction.

MODEL	RMSE VALUE	RANGE
LSTM	0.38	(0.2 - 0.5)
	0.35	(0.2 - 0.5)
ARIMA	0.05	(0.2 - 0.5)
	0.10	(0.2 - 0.5)

CONCLUSION

In the realm of stock market prediction, the journey of utilizing Long Short-Term Memory (LSTM) for forecasting closing prices has yielded valuable insights. The project's conclusion underscores the potential of LSTM in capturing intricate temporal dependencies within financial data. By seamlessly integrating machine learning with the complexities of stock markets, the project has demonstrated the efficacy of LSTM in generating accurate predictions.

The accomplishments stand as a testament to the power of modern techniques in providing actionable insights for investors and traders. The LSTM model's ability to comprehend patterns and trends in historical data, outperform baseline models, and adapt to real-time market fluctuations showcases its relevance in decision-making. While the path was marked with challenges, they illuminated the significance of data preprocessing, model fine-tuning, and real-world application.

In essence, this project signifies a step toward leveraging technology to navigate the uncertainties of financial markets. As the conclusion is drawn, the door opens to a future where data-driven approaches play a pivotal role in enhancing strategic decision-making, ultimately fostering more informed and effective financial endeavors.

REFERENCES

- Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, *167*, 599-606.
- Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, *9*, 100320
- https://www.analyticsvidhya.com/blog/2021/10/machine-learning-for-stock-market-prediction-with-step-by-step-implementation/
- https://www.analyticsvidhya.com/blog/2021/07/stock-market-forecasting-using-time-series-analysis-with-arima-model/
- https://intellipaat.com/blog/what-is-lstm/#:~:text='%20LSTM%20stands%20for%20long%20short,especially%20in%20sequence%20prediction%20problems
- https://youtu.be/LfnrRPFhkuY.
- https://analyticsindiamag.com/comparing-arima-model-and-lstm-rnn-model-in-time-series-forecasting/
- https://www.youtube.com/watch?v=CbTU92pbDKw&t=296s&pp=ygUbbHN0bSBzdG9jayBwcmljZSBwcmVkaWN0aW9u