



# Research Project Progress Report

## Week [9] – SIT723

<b>Student Name:</b>	Satvik Sharma
<b>Supervisors' Names:</b>	Prof. Maia Angelova
<b>Project Title:</b>	Financial Time Series Analysis with Machine Learning
<b>SIT723 Target Grade:</b>	Distinction I believe I have completed the tasks according to the deadline I chose. I have completed the literature review and have started working on the Project design.
<b>Overleaf Project Link:</b>	<a href="https://www.overleaf.com/read/frvfvmnnrstb">https://www.overleaf.com/read/frvfvmnnrstb</a>
<b>Project Folder Link:</b>	<a href="https://deakin365.sharepoint.com/:f:/r/sites/SIT723-FinancialTimeSeriesAnalysiswithMachineLearning/Shared%20Documents/General/Satvik%20Sharma(218595095)?csf=1&amp;web=1&amp;e=laxNjw">https://deakin365.sharepoint.com/:f:/r/sites/SIT723-FinancialTimeSeriesAnalysiswithMachineLearning/Shared%20Documents/General/Satvik%20Sharma(218595095)?csf=1&amp;web=1&amp;e=laxNjw</a>
<b>Worklog:</b>	242 hours  <a href="https://deakin365.sharepoint.com/:x:/r/sites/SIT723-FinancialTimeSeriesAnalysiswithMachineLearning/Shared%20Documents/General/Satvik%20Sharma(218595095)/Worklog.xlsx?d=w2cfbd3ab588d42f58966474ef14f85c2&amp;csf=1&amp;web=1&amp;e=2nMoH6">https://deakin365.sharepoint.com/:x:/r/sites/SIT723-FinancialTimeSeriesAnalysiswithMachineLearning/Shared%20Documents/General/Satvik%20Sharma(218595095)/Worklog.xlsx?d=w2cfbd3ab588d42f58966474ef14f85c2&amp;csf=1&amp;web=1&amp;e=2nMoH6</a>
<b>Project Plan</b>	



<b>Summary of the work planned with your supervisor:</b>	<p>In the first four weeks, the first and foremost task was to complete the Literature Review. The advantage of doing the literature review was to understand what kind of machine learning models are used by the people by the researchers.</p> <p>Next step was to create a research design and start with the machine learning project. Choose a machine learning model and evaluate stocks with the help of the chosen machine learning model.</p>																												
<b>Summary of the work done:</b>	<p>The literature review has been completed and a number of papers have been reviewed. The next part, that is, completing of the research design has been started along with the creation of the artefact. The machine learning model will be created in the Jupyter notebook. The data pre-processing has been started and the machine learning model will be made. Research on different machine learning algorithms is still in process and the best machine learning model(s) will be chosen for the project. For this task I chose a dataset of 10 years of Nifty-50 stock data and worked on the return value of the stock. The return value for me is defined as the previous date closing value subtracting from closing value divided by previous closing value. The models which I chose to work on were all statistical models and belonged from the ARIMA family. throughout the next five weeks I worked on developing them and correcting them.</p> <p>I have also worked out on evaluating the models using the mean absolute error, <math>R^2</math>, and root mean squared error. The comparison table for different algorithms implemented is given below. The stock chosen for this table is HDFC, which is one of the top stocks in NIFTY-50. The graphs of the predictions are given below</p> <table><tr><td></td><td><i>MAE</i></td><td><i>R<sup>2</sup></i></td><td><i>RMSE</i></td></tr><tr><td><i>AR</i></td><td><i>1.297</i></td><td><i>-0.033</i></td><td><i>1.738</i></td></tr><tr><td><i>ARIMA</i></td><td><i>1.299</i></td><td><i>-0.036</i></td><td><i>1.7412</i></td></tr><tr><td><i>ARIMAX</i></td><td><i>0.415</i></td><td><i>0.89</i></td><td><i>0.589</i></td></tr><tr><td><i>SARIMA</i></td><td><i>1.307</i></td><td><i>-0.05</i></td><td><i>1.752</i></td></tr><tr><td><i>SARIMAX</i></td><td><i>0.812</i></td><td><i>0.814</i></td><td><i>1.18</i></td></tr><tr><td><i>HWES</i></td><td><i>1.309</i></td><td><i>-0.03</i></td><td><i>1.74</i></td></tr></table>		<i>MAE</i>	<i>R<sup>2</sup></i>	<i>RMSE</i>	<i>AR</i>	<i>1.297</i>	<i>-0.033</i>	<i>1.738</i>	<i>ARIMA</i>	<i>1.299</i>	<i>-0.036</i>	<i>1.7412</i>	<i>ARIMAX</i>	<i>0.415</i>	<i>0.89</i>	<i>0.589</i>	<i>SARIMA</i>	<i>1.307</i>	<i>-0.05</i>	<i>1.752</i>	<i>SARIMAX</i>	<i>0.812</i>	<i>0.814</i>	<i>1.18</i>	<i>HWES</i>	<i>1.309</i>	<i>-0.03</i>	<i>1.74</i>
	<i>MAE</i>	<i>R<sup>2</sup></i>	<i>RMSE</i>																										
<i>AR</i>	<i>1.297</i>	<i>-0.033</i>	<i>1.738</i>																										
<i>ARIMA</i>	<i>1.299</i>	<i>-0.036</i>	<i>1.7412</i>																										
<i>ARIMAX</i>	<i>0.415</i>	<i>0.89</i>	<i>0.589</i>																										
<i>SARIMA</i>	<i>1.307</i>	<i>-0.05</i>	<i>1.752</i>																										
<i>SARIMAX</i>	<i>0.812</i>	<i>0.814</i>	<i>1.18</i>																										
<i>HWES</i>	<i>1.309</i>	<i>-0.03</i>	<i>1.74</i>																										



<b>Next steps:</b>	The next step is to finish up the model and start with the research article completion on the Overleaf and submit with the proper forecasting for at least one year or six months depending on how accurate the model is. Furthermore, the next step could be developing the hybrid machine learning model in the next iteration of the project, that is in the second part of the project. Including that, what I really want to do is like create portfolios for different people based on their duration they want to invest their money.
<b>Overall project progress:</b>	The project is going on with the good speed and with the current progress the project will be completed on time.

## Reflection against Assessment Rubric

Add comments and evidence against your target grade for each criterion in the assessment rubric. You should also provide evidence (relevant to your target grade) on how you meet all the relevant criteria for that grade based on your progress so far.

*In case you have not started work on a specific criterion, e.g., research evaluation, leave it blank.*

Your reflection on work done so far against the assessment rubric criteria	
<b>Literature Review</b>	High Distinction
<b>Technical &amp; Academic Writing</b>	High Distinction
<b>Research Design</b>	Credit
<b>Project Management</b>	High Distinction
<b>Artefact Development</b>	Distinction
<b>Research Evaluation</b>	Credit
<b>Research Dissemination</b>	Credit



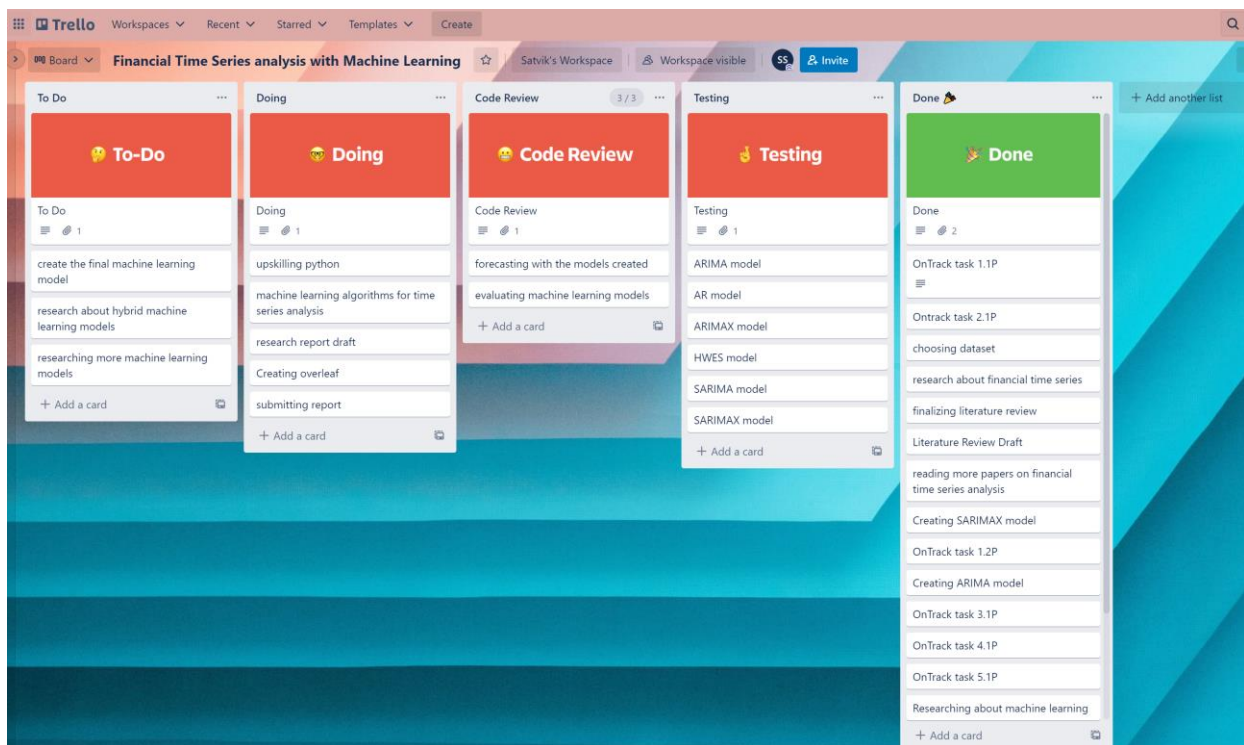
OVERALL	Distinction
---------	-------------

## Additional Evidence

Add any evidence that supports your case for your target grade. For example, screenshot of git commits log, screenshot of your cloud storage folder, progress in your research report or artefact development, or a visual chart to illustrate project progress so far.

*This is your research project, and you need to demonstrate (**with evidence**) how you meet criteria for a certain grade. Make this project report easy to read and help anyone reading this report assess, with least effort, the progress you have made.*

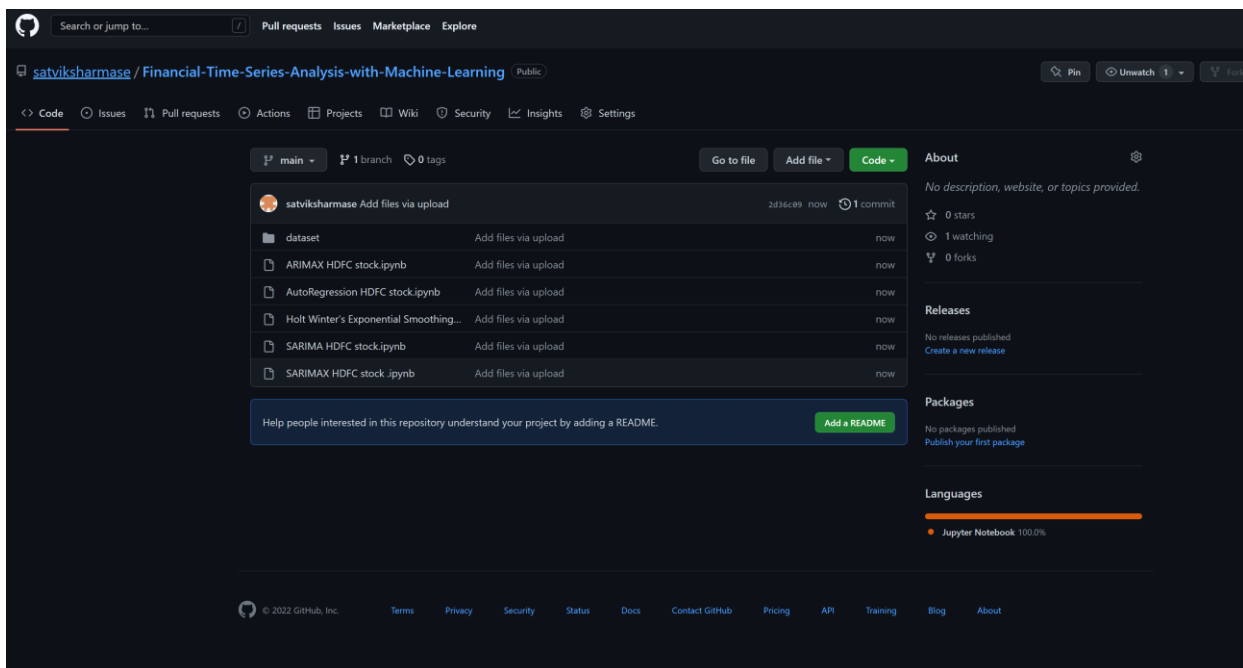
### Trello Board





## GitHub

<https://github.com/satviksharmase/Financial-Time-Series-Analysis-with-Machine-Learning>



## Literature review



used in comparison with the Logical regression, SVM, and Denoising Autoencoders and it was found that the other machine learning models performed better than ARIMA because the other machine learning models took external factors into account as well[4]. One of the papers, compares the performance of LSTM and ARIMA, and in conclusion it was found that LSTM based models run much better than the ARIMA based models even if the dataset was for one month [5].

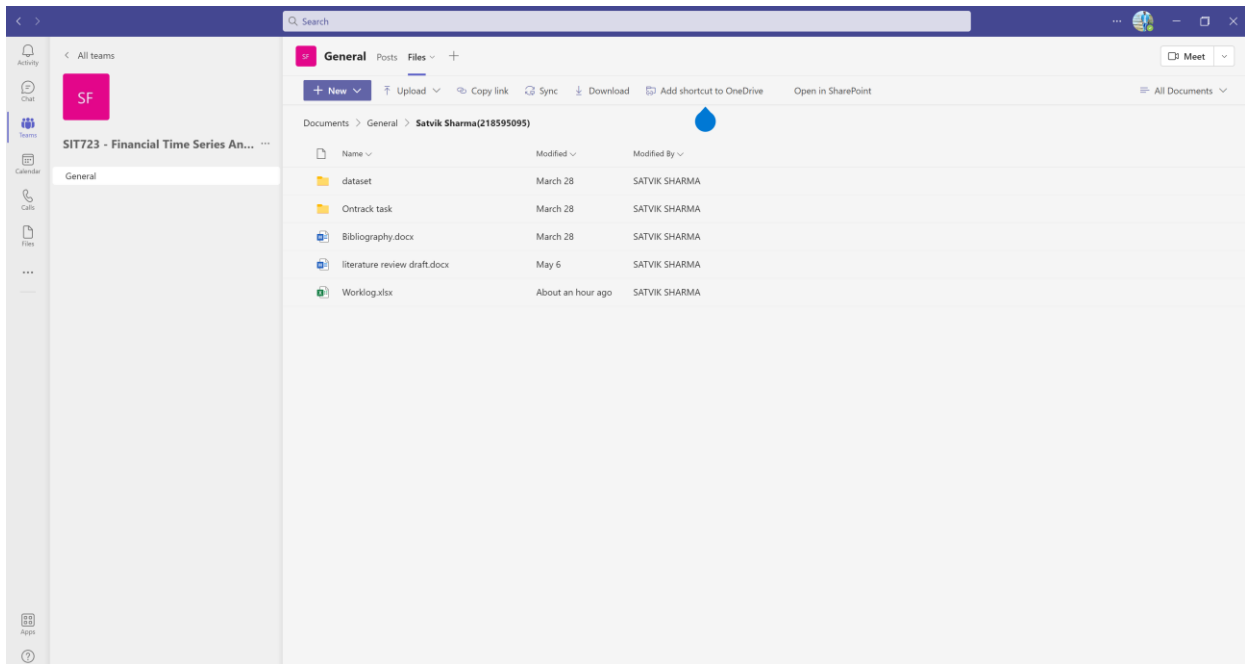
There are number of papers which have been mentioned in the table below, that have been reviewed and thus analysed what machine learning algorithms outperforms others. The table will show a small summary of what machine learning models were used and what was the conclusion of the papers

Model Name	Paper	Performance criteria	feature dataset	Time period	environment	conclusion
ARIMA, LR, MLP, SVM, DAE	[4]	-	S&P, Nasdaq, Dow 30	5-years (2012-2016)	Python Statsmodels, Sklearn, Theano	Machine learning models perform better than ARIMA and SVM outperforms all other models
ARIMA, LSTM	[5]	% Reduction in RMSE	Dow Jones Index	1-month	Python with Keras and Theano	LSTM based models work better than ARIMA
SVM, ANN, KNN	[6]	RMSE	DAX 30, S&P500	10-years (2004-2014)	-	KNN outperforms other models
KNN, SVM, Gaussian process, MLP	[7]	-	M3 time series competition data	Monthly for thousand time series	-	Best two models are MLP and gaussian progression
ANN, ARMAX, 3-D Hydrodynamic model	[8]	MAE, RMSE, R	Water level fluctuation in alpine lake	1-year (2009-2010)	Python Statsmodels	The best working models are ANN and ARIMAX
ARIMA, ANN, LSTM	[9]	P values, graphical observations	Dell's stock price	1 year (2010)	-	ANN model is better than ARIMA model and LSTM, but the hybrid of ARIMA-GARCH model can be used for more accuracy.
ANN, LSTM	[10]	nRMSE, MAPE, R <sup>2</sup>	Prediction of Solar irradiance	Hourly	Google Colab, Python scikit-learn, Keras	Due to small dataset, the ANN model works better than the LSTM, but both works much better than the persistence model
ARIMA, GARCH, regime-switching	[11]	RMSE	Home price indices by the OFHEO	20-years (1980-2000)	-	Regime switching models perform better than ARIMA and GARCH.

ARIMA-ANN, ARIMA-Kalman	[12]	MAE, MAPE, MSE	Wind speed prediction	-	-	Both hybrid models have good forecasting accuracy and suitable for wind sampling.
ARIMA-SVM, ARIMA-ANN, ARIMA-Random Forest	[13]	-	Indian stock trend	5-years (2004-2009)	MATLAB6.1, SPSS13.0	The hybrid model ANN, ARIMA, was able to predict great values than other models
AR, ARDL, KNN, SVR, Naive, VAR, MLP	[14]	RMSE, R <sup>2</sup>	Inflation forecasting	30-years (1984-2014)	-	SVR and ARDL outperforms other models and machine learning models work best with more volatile and irregular series.
Hybrid ARIMA models	[15]	-	Canadian lynx time series, sunspot time series, airline and star data	Different time periods	-	The hybrid system leads to a higher accuracy in prediction
ANN, KNN	[16]	-	Recorded EEG signals	Patient data set	-	ANN classifier's accuracy and sensitivity was higher than that of KNN classifier.
ANN, LSTM, MLR	[17]	MSE, r, RMSE	Prediction of irrigation groundwater quality	-	Python	ANN and MLR model have highest accuracy in multiple scenarios
LR, LSTM	[18]	RMSE, MAPE	For hire vehicles and yellow taxi	-	-	LR is used to select the important variables and LSTM helps to improve the accuracy.
ARIMA	[19]	-	New York Stock exchange and Nigeria stock exchange	-	python	ARIMA has strong potential for short term prediction
MLR, KNN, ANN, ANFIS	[20]	Mash-Sutcliffe coefficient	Stream flow prediction	Monthly	-	The accuracy of each of the model depends on the condition, but the hybridisation was effective.

After observing the table above, it can be clearly seen that the ARIMA is not a good option for the analysis of financial time series even though it is one of the most common methods used. Multiple machine learning methods like LSTM and ANN are also most commonly used as well as have good

## Project folder



## Worklog



Worklog.xlsx									
Search									
File Home Insert Draw Page Layout Formulas Data Review View Automate Help Open in Desktop App Tell me what you want to do Editing Copy									
Calibri 12 B Conditional Formatting Styles Format									
G165 X B implementing the machine learning algorithms									
108	Satish Sharma	Musa Angelina	Financial Time Series Analysis	25-04-2022	120	Project Management	managing the project with the help of Jira and the online repositories		
109	Satish Sharma	Musa Angelina	Financial Time Series Analysis	25-04-2022	120	OnTrack Task	completing the OnTrack task for the sixth week		
110	Satish Sharma	Musa Angelina	Financial Time Series Analysis	27-04-2022	60	Research Design	researching on different machine learning algorithms		
111	Satish Sharma	Musa Angelina	Financial Time Series Analysis	27-04-2022	60	Supervision Meeting	meeting with the supervisor to discuss the next part of the project		
112	Satish Sharma	Musa Angelina	Financial Time Series Analysis	28-04-2022	90	Up-skilling	up-skilling on how to create the machine learning models		
113	Satish Sharma	Musa Angelina	Financial Time Series Analysis	28-04-2022	90	Research Design	researching on different machine learning algorithms especially SARIMAX		
114	Satish Sharma	Musa Angelina	Financial Time Series Analysis	28-04-2022	60	Artifact Implementation	implementing and fixing the errors in the code		
115	Satish Sharma	Musa Angelina	Financial Time Series Analysis	28-04-2022	120	Up-skilling	up-skilling on how to create the machine learning models		
116	Satish Sharma	Musa Angelina	Financial Time Series Analysis	29-04-2022	90	Writing	editing the literature review		
117	Satish Sharma	Musa Angelina	Financial Time Series Analysis	29-04-2022	90	Research Design	researching new and better machine learning algorithms		
118	Satish Sharma	Musa Angelina	Financial Time Series Analysis	29-04-2022	60	Artifact Implementation	implementing and completing the project design		
119	Satish Sharma	Musa Angelina	Financial Time Series Analysis	30-04-2022	90	Up-skilling	upskilling on how to use LaTeX		
120	Satish Sharma	Musa Angelina	Financial Time Series Analysis	30-04-2022	60	Artifact Implementation	implementing the machine learning algorithms		
121	Satish Sharma	Musa Angelina	Financial Time Series Analysis	01-05-2022	120	Research Design	research about the machine learning algorithms that can be used for time series analysis		
122	Satish Sharma	Musa Angelina	Financial Time Series Analysis	01-05-2022	60	Artifact Implementation	implementing the machine learning algorithms		
123	Satish Sharma	Musa Angelina	Financial Time Series Analysis	01-05-2022	120	Literature Review	updating the Literature Review		
124	Satish Sharma	Musa Angelina	Financial Time Series Analysis	02-05-2022	90	Project Management	managing the project with the help of Jira and the online repositories		
125	Satish Sharma	Musa Angelina	Financial Time Series Analysis	02-05-2022	60	OnTrack Task	fixing the OnTrack task for the sixth week		
126	Satish Sharma	Musa Angelina	Financial Time Series Analysis	02-05-2022	60	OnTrack Task	completing the OnTrack task for the seventh week		
127	Satish Sharma	Musa Angelina	Financial Time Series Analysis	03-05-2022	120	Research Design	researching new and better machine learning algorithms		
128	Satish Sharma	Musa Angelina	Financial Time Series Analysis	03-05-2022	120	Up-skilling	upskilling machine learning by watching youtube videos and tutorials		
129	Satish Sharma	Musa Angelina	Financial Time Series Analysis	04-05-2022	90	Research Design	researching the papers that can be used for machine learning		
130	Satish Sharma	Musa Angelina	Financial Time Series Analysis	04-05-2022	120	Artifact Implementation	implementing and completing the project design		
131	Satish Sharma	Musa Angelina	Financial Time Series Analysis	04-05-2022	60	Supervision Meeting	meeting with the supervisor to discuss the next part of the project		
132	Satish Sharma	Musa Angelina	Financial Time Series Analysis	04-05-2022	90	Up-skilling	upskilling machine learning by watching youtube videos and tutorials		
133	Satish Sharma	Musa Angelina	Financial Time Series Analysis	04-05-2022	120	Research Design	researching on different machine learning algorithms		
134	Satish Sharma	Musa Angelina	Financial Time Series Analysis	04-05-2022	60	Artifact Implementation	implementing and fixing the errors in the code		
135	Satish Sharma	Musa Angelina	Financial Time Series Analysis	07-05-2022	60	Literature Search	searching different algorithms for the project		
136	Satish Sharma	Musa Angelina	Financial Time Series Analysis	07-05-2022	120	Up-skilling	upskilling with machine learning in python		
137	Satish Sharma	Musa Angelina	Financial Time Series Analysis	08-05-2022	120	Research Design	researching on different machine learning algorithms		
138	Satish Sharma	Musa Angelina	Financial Time Series Analysis	08-05-2022	120	Artifact Implementation	implementing the machine learning algorithms		
139	Satish Sharma	Musa Angelina	Financial Time Series Analysis	09-05-2022	120	Project Management	managing the project with the help of Jira and the online repositories		
140	Satish Sharma	Musa Angelina	Financial Time Series Analysis	09-05-2022	60	Writing	writing the latex		
141	Satish Sharma	Musa Angelina	Financial Time Series Analysis	09-05-2022	60	OnTrack Task	completing the OnTrack task for the eighth week		
142	Satish Sharma	Musa Angelina	Financial Time Series Analysis	10-05-2022	60	Research Design	researching on different machine learning algorithms		
143	Satish Sharma	Musa Angelina	Financial Time Series Analysis	10-05-2022	60	Up-skilling	upskilling machine learning by watching youtube videos and tutorials		
144	Satish Sharma	Musa Angelina	Financial Time Series Analysis	10-05-2022	120	Other	writing the research artifact draft		
145	Satish Sharma	Musa Angelina	Financial Time Series Analysis	11-05-2022	60	Research Design	researching on different machine learning algorithms from ARIMA family		
146	Satish Sharma	Musa Angelina	Financial Time Series Analysis	11-05-2022	120	Up-skilling	upskilling on how to fix machine learning models		
147	Satish Sharma	Musa Angelina	Financial Time Series Analysis	11-05-2022	60	Artifact Implementation	implementing the machine learning algorithms		
148	Satish Sharma	Musa Angelina	Financial Time Series Analysis	11-05-2022	60	Supervision Meeting	meeting with the supervisor to discuss the next part of the project		
149	Satish Sharma	Musa Angelina	Financial Time Series Analysis	12-05-2022	60	Research Design	researching on correcting different machine learning algorithms implemented		
150	Satish Sharma	Musa Angelina	Financial Time Series Analysis	13-05-2022	60	Up-skilling	upskilling on how to use LaTeX		
151	Satish Sharma	Musa Angelina	Financial Time Series Analysis	13-05-2022	60	Research Design	researching the papers that can be used for machine learning		
152	Satish Sharma	Musa Angelina	Financial Time Series Analysis	13-05-2022	60	Up-skilling	upskilling with machine learning in python		
153	Satish Sharma	Musa Angelina	Financial Time Series Analysis	15-05-2022	120	Artifact Implementation	implementing and completing the project design		
154	Satish Sharma	Musa Angelina	Financial Time Series Analysis	15-05-2022	60	Research Design	researching new and better machine learning algorithms		
155	Satish Sharma	Musa Angelina	Financial Time Series Analysis	15-05-2022	60	Artifact Implementation	implementing and fixing the errors in the code		
156	Satish Sharma	Musa Angelina	Financial Time Series Analysis	15-05-2022	60	Project Management	managing the project with the help of Jira and the online repositories		
157	Satish Sharma	Musa Angelina	Financial Time Series Analysis	16-05-2022	30	Research Design	researching on correcting different machine learning algorithms implemented		
158	Satish Sharma	Musa Angelina	Financial Time Series Analysis	16-05-2022	60	Up-skilling	upskilling machine learning by watching youtube videos and tutorials		
159	Satish Sharma	Musa Angelina	Financial Time Series Analysis	16-05-2022	90	Research Design	researching the papers that can be used for machine learning		
160	Satish Sharma	Musa Angelina	Financial Time Series Analysis	18-05-2022	90	Artifact Implementation	implementing and fixing the errors in the code		
161	Satish Sharma	Musa Angelina	Financial Time Series Analysis	18-05-2022	60	Supervision Meeting	meeting with the supervisor to discuss the next part of the project		
162	Satish Sharma	Musa Angelina	Financial Time Series Analysis	19-05-2022	120	Research Design	researching in the implementation of ARIMA		
163	Satish Sharma	Musa Angelina	Financial Time Series Analysis	19-05-2022	60	Evaluation Task	evaluating the tasks done till now and checking on the progress of project		
164	Satish Sharma	Musa Angelina	Financial Time Series Analysis	20-05-2022	120	Research Design	researching on correcting the values prediction		
165	Satish Sharma	Musa Angelina	Financial Time Series Analysis	20-05-2022	60	Artifact Implementation	implementing the machine learning algorithms		
166	Satish Sharma	Musa Angelina	Financial Time Series Analysis	21-05-2022	60	Research Design	researching on correcting different machine learning algorithms implemented		
167	Satish Sharma	Musa Angelina	Financial Time Series Analysis	21-05-2022	60	Artifact Implementation	implementing and fixing the errors in the code		
168	Satish Sharma	Musa Angelina	Financial Time Series Analysis	21-05-2022	90	Evaluation Task	evaluating the implemented artifacts and checking how accurate each of the algorithms are		
169	Satish Sharma	Musa Angelina	Financial Time Series Analysis	21-05-2022	90	Other	researching on the future work for project		
170	Satish Sharma	Musa Angelina	Financial Time Series Analysis	22-05-2022	30	Project Management	managing the project with the help of Jira and the online repositories		
171	Satish Sharma	Musa Angelina	Financial Time Series Analysis	22-05-2022	120	Writing	writing the research artifact draft		
172	Satish Sharma	Musa Angelina	Financial Time Series Analysis	22-05-2022	60	OnTrack Task	completing the OnTrack task for the ninth week		
173									

Machine learning evaluation table

	MAE	$R^2$	RMSE
AR	1.297	-0.033	1.738
ARIMA	1.299	-0.036	1.7412
ARIMAX	0.415	0.89	0.589
SARIMA	1.307	-0.05	1.752
SARIMAX	0.812	0.814	1.18
HWES	1.309	-0.03	1.74



