Predictive Analysis for Earth and Mars Temperature and Pressure with ARIMA model

M Anis Mumtaz Chaudhary
x23215011
MSc Artificial Intelligence
National College of Ireland

Buse Ay
x23122633
MSc Artificial Intelligence
National Coleege of Ireland

Sachleen Singh Chani
x22244778
MSc Artificial Intelligence
National College of Ireland

***Abstract*—Humanity has always looked up to the stars for answers, where we come from and were we can go. Although there are a lot of features to consider for the prediction of weather, this paper limits the scope to temperature and pressure data. In this paper we gain insights into the weather conditions on Mars, and compare them to that of Earth. The ARIMA model is initially applied to the Earth data and subsequently to Mars data, Earth data becoming a benchmark for the analysis. The model performs excellent in detecting intricate characteristics and trends in a time-series. To train the model and display the predictions, MongoDB and MySQL have been used to store the data and a web application to display the predictions in a user-friendly way. The results from the prediction have been consistent with the data provided, and the ARIMA model was able to follow the trends accurately.**

***Keywords*—** ARIMA, Mars habitability, temperature forecast, pressure predction, MongoDB, MySQL

I. INTRODUCTION

Humanity has always been curious about the unknown, especially about what is outside of the Pale Blue Dot. They wonder about other planets and whether a new Earth might be found. Since the 1960s, this curiosity has increased day by day with various scientific missions. As we know, the potential energy sources for life are carbon and water, like Earth, Mars contains each of these in a sort of its way, and because of that Mars is chosen as the potential planet for the next generations [1]. Another reason to choose Mars for the next colonization is that Martian days are slightly more than Terrestrial days, and the two planets have similar seasons [2]. National Aeronautics and Space Administration(NASA) launched rovers like Perservance and Curiosity to explore the weather features like temperature, pressure, wind speed, and humidity for the red planet. Besides these features, these rovers and so many others are looking for various important features for habitability such as soil and water content [2]. Besides NASA, there are different space agencies also interested in Mars colonization, and the first human mission is expected soon [2]. Usage of nowadays technology such as machine learning and artificial intelligence, in space exploration expanding rapidly. While rovers collect the data, algorithms answer the questions for humanity. The first step for the Mars colonization is analyzing the weather conditions. As mentioned before Preservence and Curiosity provide daily weather data. For weather forecasting and analyses, algorithms like Artificial Neural Networks(ANNs), Recurrent Neural Networks(RNNs), and Convolutional Neurol Networks(CNNs) are used for various data analysis tasks [1]. These algorithms are good for capturing nonlinear relationships and they are more flexible. There are also models like ARIMA(AutoRegressive Integrated Moving Average) which is used for capturing linear features and Long Short Term Memory network (LSTM) which is used for capturing long-term dependencies [3]. All these different methods are useful for forecasting, they are just approaching the problem differently. Having a better understanding of Mars's atmospheric conditions and weather could be useful for handling weather anomalies for the rovers and the other research instruments, and also could help the next colonization projects. In this paper, the ARIMA model is used for forecasting Mars and Earth weather on the dataset collected from the NASA Curiosity rover and Earth Climate Assessment and Datasets. A literature review is provided in Section 2. The data cleaning process and processing activities are explained and supported with visualization in Section 3. Also, data storing tools, MongoDB and MySQL, the workflow, and the web application created by using Flask are explained in the same section. Section 3 also contains information about the machining learning model used in this project, which is the ARIMA model. In Section 4, results and evaluation are represented. And Section 5, concludes the project.

II. RELATED WORK

In the field of time-series and forecasting many studies have been done. Priyadarshini and Puri [2], explored various machine learning models to analyse the martian weather using different features into consideration. They identified that the LSTM model outperformed the rest, stacked-LSTM, GRU (Gated Recurrent Units), CNN-LSTM. These models were measured exhaustively on variety of metrics like Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R-squared coefficient and Mean Squared Error (MSE). Furthermore in the paper Priyadarshini and Puri introduced the CNN-LSTM architecture. This architecture would extract from the data through CNN and utalising the LSTM layers to predict the weather. Despite the extensive use of machine learning models, the ARIMA (Autoregressive integrated moving average) was not explored. The foundation for this ARIMA model was laid by Box and Jenkins [4], outlined in their paper. Further, even though Priyadarshini and Puri, in their paper conclude that the LSTM worked the most effective in the predictions, but doesn't explicitly states the advantages of using the LSTM model.

Cheng [5] applied the ARIMA model to gain insights into global temperature prediction, focusing particularly on earth climate change. This research shows the significance of using multiple models, in this particular case Cheng uses XGBoost-LSTM and ARIMA. Their paper predicted the global average temperature for 2050 and 2100, assuming the ecosystem remains stable. Although the approach is innovative, there are a certain limitations, the XGBoost-LSTM model is highly sensitive to the parameters chosen, easily introducing biased predictions if the parameters are not tweaked correctly.

Similarly, Guan et al. [6] in their paper focused on the prediction of the global temperature using the ARIMA model. They used the last 200 historic data points and predicted 20 future data points for the year 2019 and 2020. They concluded accurately an average increase in the global average temperature. Some limitations do exist, the short-term prediction ability for the model, introducing uncertainties while predicting extended periods.

Dimiri et al. [7] expanded the ARIMA model by including a temporal dimension by using Seasonal ARIMA (SARIMA) by taking monthly means of the temperatures and precipitation data while training the model. Even though the SARIMA model performed better, in certain extreme conditions over-predictions could be observed. Similarly, Jafarian-Namin et al [8] did a comprehensive analysis by utalilsing ARIMA and Particle Swarm Optimisation (PSO) improvement to ANN for monthly temperature data of Tehran, Iran. The challenges faced by Jafarian-Namin et al. was the integration of ANN, PSO and ARIMA for predicting the temperature.

Focusing on the mars weather predictions, Al-Saad et al. [1] in their paper proposed a CNN-LSTM integration for predicting the Mars temperature, which demonstrated better performance. Similar work was done by Pant et al. [9] in predicting the martian weather, applying the ARIMA model for the forecast. Their research highlighted the efficiency of linear regression model.

Green et al. [10] put emphasis on the importance of increased data collection and modeling from Mars to support our exploration of other planets. The paper delved into challenges and critical role of understanding the weather patterns on other planets, to gain a better understanding human survival on mars and beyond. Green et al. study forms a crucial bridge between Earth and Mars weather predictions, putting focus on a robust ”Space Weather” observation system to ensure success in exploring other planets.

In summary, though not all material discussed in this section directly relates to the mars weather prediction, but gives a detailed insight into the working and the shortcomings of the ARIMA model for forecasting.

III. Methodology

A. Data Collection and Cleaning

In this project, there are three different data sets used. One of them is taken from NASA’s Curiosity Rover [11]. On the data, there are two different date versions, Martian day(Sol), and terrestrial day. Sol is slightly longer than a terrestrial day, approximately 39 minutes [1]. Features for this data set are maximum and minimum ground and air temperature, mean pressure, sunrise and sunset times, UV radiation levels, and weather conditions which are always sunny. Also, the data set contains wind speed and humidity columns but there are no values on those columns. It starts on the 7th of August 2012 and ends on the 26th of January 2022. The dataset has 3197 rows.

For the data cleaning process on this dataset, columns are not needed and the columns that do not have any values have been removed. Decided the work with terrestrial date, the maximum and minimum air temperature, and the pressure. In the data set, there were null values like “Value not found”, these are replaced with the “NaN”. In the end, null values were less than 10%, so rows which have null values were dropped. Values were converted to the right format, dataset was divided into two, one for weather prediction which includes date, maximum and minimum temperature, and pressure, and the other one for visualization purposes which includes date, sunrise and sunset time, and the weather condition.

Earth dataset collected from Earth Climate Assessment and Datasets(ECA&D) [12], this data has daily weather records from Heathrow Airport, London. It starts on the 1st of January 1956 and ends on the 30th of November 2023. There were different files for different features, and all those dataset files were in text (.txt) format, because of that data was converted to a data frame for every feature separately. For context, the pressure data file has the information about the date, pressure, and quality of data, for maximum and minimum temperature was the same issue. The data quality column has three different values “0” means data quality is good, “9” means missing value, and “1” means unknown data quality. After the data is converted to the right format, merged into one data frame. Only data assigned as “0” is kept and the rest is dropped.

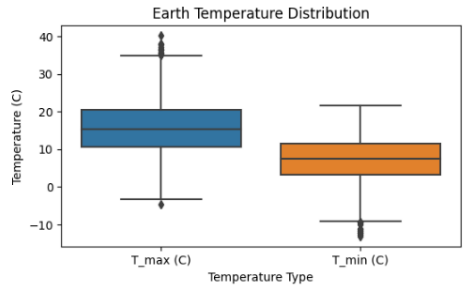

Fig. 1. Earth Min and Max temp(C) distribution

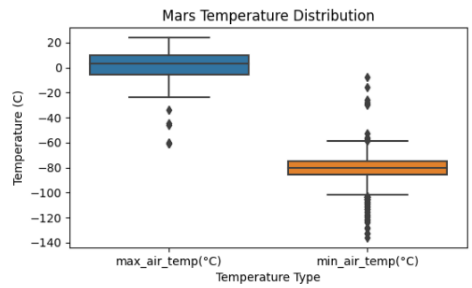

Fig. 2. Mars Min and Max temp(C) distribution

As you can see in Figure 1, Figure 2, Earth, and Mars temperature distributions and in Figure 3 and Figure 4 correlation matrices show that the mean maximum and minimum temperatures for Earth are close to each other so, the correlation is higher.

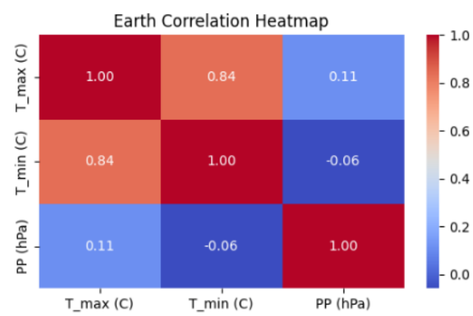

Fig. 3. Earth Pearson’s Correlation

But for Mars, the situation is different. Mars weather is not suitable for humans, less correlation between the temperature comparing the Earth can be the result of that.

Fig 5 and Fig 6 shows the comparison of maximum temperature and pressure.

Fig 7 shows the scatter plot of temperature vs pressure. Earth data, as seen in all graphs, is consistent. Mars data is also consistent but not as much as Earth. There are a couple of outliers, these might be the results of extreme Mars weather conditions. They kept it as it was to get a better overall look.

Finally, we used the python’s Astral library [13] to generate the sunset and sunrise times for earth for the particular dates we gathered the temperature and pressure data.

These two datasets, out of which the temperature and pressure data was stored into MongoDB to

B. DataBases - MongoDB-MySQL

In this project, two different databases are used, MongoDB and MySQL.

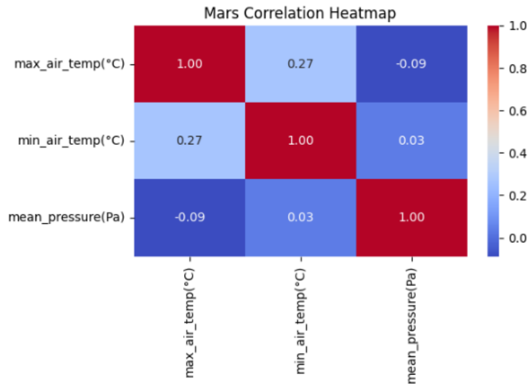

Fig. 4. Mars Pearson's Correlation

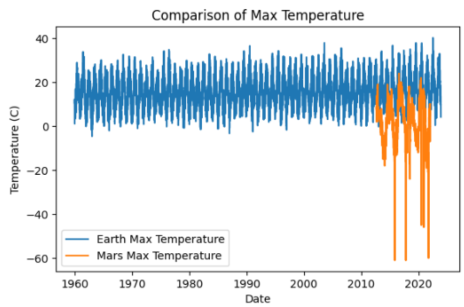

Fig. 5. max temperature of Earth and Mars

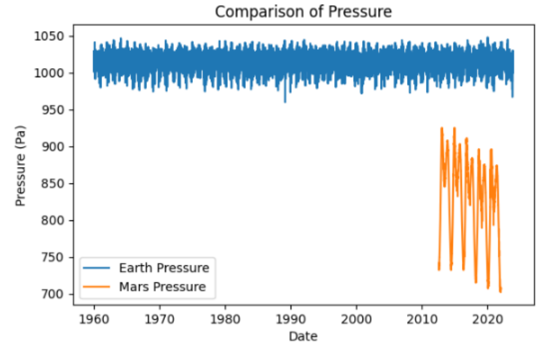

Fig. 6. Pressure comparison of Earth and Mars

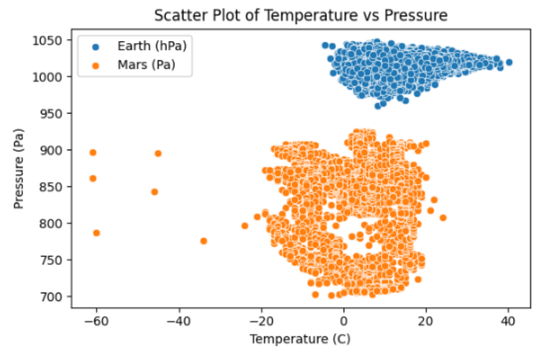

Fig. 7. Scatter plot of max temp and pressure of Earth and Mars

MongoDB is a NoSQL database system that uses JSON documents and organizes them into tables and rows. It is flexible and can handle unstructured and semi-structured data. MongoDB is popular for web applications, management systems, and analytics.

MySQL is open source, SQL database management system. It needs a predefined structure and it organizes the data into rows and columns. Both are effective database systems, but MySQL needs a structured database model and MongoDB uses a document-based database model for unstructured or semi-structured.

In this project, as you can see in Fig 8, after data preprocessing, all datasets are stored in MongoDB, and then Temperature and pressure datasets are used in the training model by fetching them from MongoDB. After the training structured results, are stored in MySQL. For the web application, data which is stored in the MySQL is used.

The MySQL database was structured to capture the forecasted temperature and pressure data for both the Earth and Mars. A dedicated table was used to store each of the planets data.

Overall Database Workflow:

- Data Integration: The clean historic data for both Earth and Mars was stored into MongoDB. Subsequently this data becomes the primary source for training the ARIMA model.
- ARIMA training: The data extracted from MongoDB using python's MongoDB is used to train the model. The predicted future pressure and temperature is stored into MySQL.
- MySQL: The forecast data is stored into two separate tables with the dates.
- Web App: The data stored in MySQL is retrieved to be used to display in the web application. While the historic data is taken from MongoDB.

C. ARIMA Model

Autoregressive Integrated Moving Average (ARIMA) model is a flexible time-series model to predict the future values based on the past data given to the model. For this particular paper, we trained the model using the historic minimum and maximum temperature for Earth and Mars, and the historic pressure data for both the planets.

ARIMA model is a versatile methods, able to capture patterns in the time-series data given to for training. This robustness comes from the model to be able to combine three fundamental components, Autoregression (AR), Integration (I) and finally Moving Average (MA). This approach ensure that the model can identify complex time-series details [14].

To further elaborate these components that make up the ARIMA model:

Autoregression: This component of the model focuses on the current data point and the preceding values. By doing this, the model can form a trend from the data given to it.

Integration: While this component tackles the issue of non-stationarity. it does this by subtracting the current value from the previous value point in the time-series.

Moving Average: This component of the model acts as a bridge, connecting the current data point that is in focus and the residual error that is generated due to the moving average of the past data points.

Each of these components are explicitly depicted in the model code as the parameters as p, d and q. This forms the standard notation for the model, i.e. ARIMA(p,d,q). These parameters are integer values, which need to be tweaked so the model works with the time-series, which in this case is the weather information of earth and mars. Fine-tuning these parameters become crucial to align the model with the unique and intricate features present in out data and subsequently enhance the forecasting ability.

D. Web Application - Flask

The results, i.e. the predicted values from the model we stored in MySQL database, creating two separate relations for Earth prediction

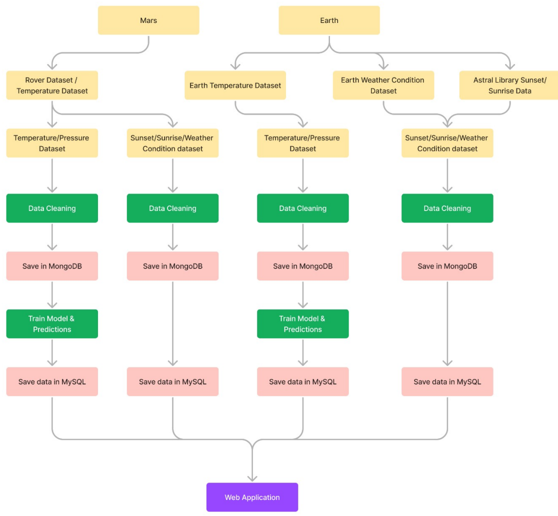

Fig. 8. Flowchart

and mars prediction of temperature (min and max) and pressure. Using Flask we created a user friendly web application to display this predictive data.

The Web application is interactive, the users can input the date for which they want to observe the predictions. The website would give the output of the minimum temperature, maximum temperature and pressure. Along with this information, the sunset sunrise times and the weather conditions as well.

IV. Results and Evaluation

The historic data for both the Earth data and Mars data show distinctive trends, which the model was able to predict.

The ARIMA model prediction for the temperature and pressure data, the forecast generated were stored in MySQL.

The step for the model was set at 60, i.e. 60 days (2 months). This value can be increased to predict further.

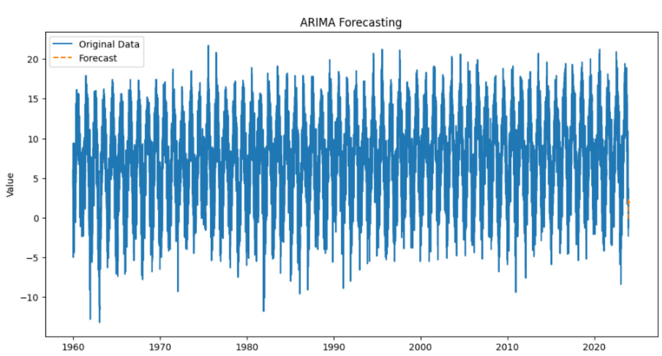

Fig. 9. Historic(blue) and predicted(orange) minimum temperature of Earth

Fig 9, we can see the prediction plotted along with the historic minimum temperature data for earth. As we can observe the temperature data is quite consistent, staying withing a stable range. We can observe that the predicted temperature follows the trend closely.

Similarly, we can observe in Fig 10 that the predicted values for the minimum temperature for mars is consistent with the historic

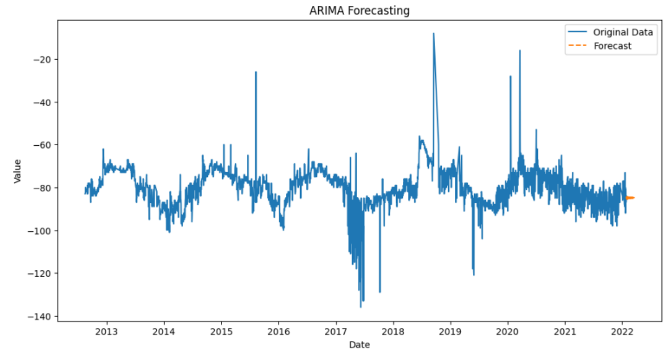

Fig. 10. Historic(blue) and predicted(orange) minimum temperature of Mars

data. Predicting the data using Mar's extreme conditions becomes a challenge for the ARIMA model, as compared to the much stable Earth data.

The web app played an instrumental part in viewing these predicted data in a clear and concise way for users. Giving the users the minimum and maximum temperature and pressure for Earth and Mars, along with the sunset and sunrise times and weather conditions.

V. Conclusions and Future Work

In this paper we outlined the used of the time-series prediction of Earth and Mars weather selected features, minimum and maximum temperature and pressure. The model was able to accurately follow the trend in the historic data provided for training. This gives us a comparison between the stable, habitable conditions on Earth with the harsh, constantly fluctuating conditions on Mars.

The thinner atmosphere of Mars makes cases these extreme conditions and with the lack of a magnetic field, the temperatures are at the whim of the Sun, the side facing the sun reaching a warm 20 degree Celsius and the side facing away as low as -153 degree Celsius. As we gaze towards the stars for the future of humanity, this research provides a rudimentary basis for the selection of a possible next home for us.

For future work, we can include more features like wind speeds, UV radiation, etc. and used a hybrid model like ARIMA and LSTM to get more accurate predictions.

References

[1] M. Al-Saad, N. Aburaed, S. Al Mansoori, W. Mansoor, and H. Al-Ahmad, "A study on deep learning approaches for mars weather forecasting," in *2022 5th International Conference on Signal Processing and Information Security (ICSPIS)*, 2022, pp. 76–79.

[2] I. Priyadarshini and V. Puri, "Mars weather data analysis using machine learning techniques," *Earth Science Informatics*, vol. 14, no. 4, pp. 1885–1898, December 1 2021. [Online]. Available: https://doi.org/10.1007/s12145-021-00643-0

[3] L. Ji, Y. Zou, K. He, and B. Zhu, "Carbon futures price forecasting based with arima-cnn-lstm model," *Procedia Computer Science*, vol. 162, pp. 33–38, 01 2019.

[4] G. E. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970.

[5] X. Cheng, "Global temperature study based on arima and xgboost-lstm models," *Frontiers in Economics and Management*, vol. 4, no. 4, pp. 224–234, Apr 2023.

[6] G. Naiqian, G. Yuxin, and S. Xuelian, "Global temperature forecast based on arima model," in *2019 4th International Conference on Communication and Information Systems (ICCIS)*, 2019, pp. 108–112.

[7] T. Dimri, S. Ahmad, and M. Sharif, "Time series analysis of climate variables using seasonal arima approach," *Journal of Earth System Science*, vol. 129, no. 1, p. 149, June 27 2020. [Online]. Available: https://doi.org/10.1007/s12040-020-01408-x

[8] S. Jafarian-Namin, D. Shishebori, and A. Goli, “Analyzing and predicting the monthly temperature of tehran using arima model, artificial neural network, and its improved variant,” *Journal of Applied Research on Industrial Engineering*, pp. –, 2023. [Online]. Available: https://www.journal-aprie.com/article_167429.html
[9] P. Pant, A. S. Rajawat, S. B. Goyal, B. B. Kemat, T. C. Mihălţan, C. Verma, and M. S. Răboacă, “Machine learning techniques for analysis of mars weather data,” in *2023 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2023, pp. 1–7.
[10] J. L. Green, C. Dong, M. Hesse, C. A. Young, and V. Airapetian, “Space weather observations, modeling, and alerts in support of human exploration of mars,” *Frontiers in Astronomy and Space Sciences*, vol. 9, Nov. 2022. [Online]. Available: http://dx.doi.org/10.3389/fspas.2022.1023305
[11] DeepContractor, “Mars rover environmental monitoring station,” https://www.kaggle.com/datasets/deepcontractor/mars-rover-environmental-monitoring-station, Year Accessed, accessed on: 6 Jan 2024.
[12] “European climate assessment & dataset (ecad),” https://www.ecad.eu/, accessed on: 6 Jan 2024.
[13] T. A. Project, “Astral: Python calculations for the position of the sun and moon,” https://pypi.org/project/astral/, Year, accessed: 4 Jan 2024.
[14] W. Blanzeisky. (2020) Predicting temperature data using arima model: Time series. Accessed: 6 Jan 2024. [Online]. Available: https://medium.com/@williamblanzeisky/predicting-temperature-data-using-arima-model-time-series-b85eb6a754cb