To: David Lindsay  
From: Henry Forget, Inessa Khodak, Nico Loreto, Hadiyah Kahn, Kiet Tran  
Date: Thursday, 6th, 2024  
Subject: Needs Assessment Assignment

**Purpose Statement**

Two machine learning models will be proposed to Ideabytes for two separate business objectives. The first will be chosen to be the best-performing machine learning algorithm to predict temperature values over time intervals of 1 hour, 4 hours, and 8 hours, based on time-series data obtained from IoT devices located in industrial freezers. The second algorithm is for detecting anomalies based on parameters like vibration, sound, and electricity data from IoT devices located on Industrial Compressors. Various machine learning algorithms will be tested on sample data sets. Based on the efficiency and accuracy of the obtained results from the tests, the two best machine learning models will be prepared, trained on historical data and integrated into Ideabytes Cloud environment. Once integrated, Ideabytes’ clients will be alerted in the event sampled values cross established thresholds.

**Problem Statement**

Ideabytes offers its clients IoT solutions that measure and log various parameters like temperature, humidity, vibration, pressure, etc. To monitor and visualize such data Ideabytes provide web dashboards and Mobile apps. However, for clients to better manage their processes, Ideabytes needs to provide a way to predict certain events, such as increases in temperature and anomalous vibration. Lack of such service can lead to spoiling products worth millions of dollars and malfunction of important equipment. Although real-time dashboards help clients follow data trends in how equipment performs, they are not designed to plan for preemptive actions. Machine learning models that can predict future values and detect performance anomalies with a good level of accuracy will help clients make informed decisions prior to machine failures and lower the risk of damaged merchandise.

**Project Criteria**

To be considered a success, chosen machine learning models:

·        Should have accuracy above 70%.

·        Should have recall above 70%.

·        Should be able to predict temperature after 1 hour, 4 hours, and 8 hours for freezer storage.

·        Should be able to detect anomalies for machine vibration, i.e.:

a)      Identify the anomaly

b)     Detect outside of  range vibration values

c)     Detect vibration trending to higher values

·        Should be designed to integrate with the client’s monitoring and alerting system.

**Investigation**

Research

1. IdeaBytes will be supplying us with 6 years worth of data over several machines to train the chosen machine learning algorithms.
2. Several online sources have given guidance on everything we need to consider for our specific circumstances. Here are some examples of sources we have found useful in guiding us towards our goal:
   1. Our project will use Python as a coding language. Preliminary learning in both the language itself and important data analysis libraries is using Python for Data Science - Course for Beginners [1].
   2. We learnt about several different types of machine learning algorithms as well as their strengths and weaknesses to choose the two that best suit our needs as well as what libraries to use to best simulate said algorithms [2], [3], [4], [5].
   3. We found several sources that will guide us in the implementation of machine learning models in Python [6], [4].

Tentative Findings

Based on the research done thus far, the models we are leaning towards are as follows:

* ARIMA for the purpose of trend detection as it facilitates predicting future outcomes based on a historical time series as well as being one of the two most widely used approaches to data prediction.
* Isolation forest model for the purpose of anomaly detection as it generates anomaly scores between 0 and 1, based on a point’s location relative to the other points. Using this we can account for the aforementioned external influences by tweaking its tolerance. This model is also well documented and is included in the free machine learning library for Python, scikit-learn.

**Future Plans**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Task Description** | **Task Duration** | **Due Date** | **People Responsible** |
| 1 | Familiarize ourselves with python and it's associated ML libraries | 2 weeks | 07/06/2024 | All |
| 2 | Research the processes and algorithms needed for both sections of the project and divide the team | 1 week | 14/06/2024 | All |
| 3 | Both teams create their proof of concept and present it to the client, clarifying any issues present | 3 weeks | 05/07/2024 | Both groups individually |
| 4 | Plan out the system structure for integration with their existing API’s for getting and giving data | 4 weeks | 02/08/2024 | All |
| 5 | Both teams complete their ML processes with all criteria met | 10 weeks | 11/10/2024 | Both groups individually |
| 6 | Integrate our ML programs with the IdeaBytes system, tweaking any issues if needed | 5 weeks | 15/11/2024 | All |
| 7 | Complete the documentation and related presentations | 3 weeks | 06/12/2024 | All |

**Citations**

[1] freeCodeCamp.org. Python for Data Science - Course for Beginners (Learn Python, Pandas, NumPy, Matplotlib). (Jun 2, 2020). Accessed: June 4th, 2024. [Online Video]. Available: https://www.youtube.com/watch?v=LHBE6Q9XlzI

[2] Simplilearn. Pytorch vs TensorFlow vs Keras | Which is Better | Deep Learning Frameworks Comparison | Simplilearn. (Jan 30, 2021). Accessed: June 4th, 2024. [Online Video]. Available: https://www.youtube.com/watch?v=4L86D\_fU6sQ

[3] N. Selvaraj, “8 machine learning models explained in 20 minutes,” DataCamp, https://www.datacamp.com/blog/machine-learning-models-explained (accessed Jun. 4, 2024).

[4] A. R. Kulkarni, A. Shivananda, A. Kulkarni, and V. A. Krishnan, Time series algorithms recipes: Implement Machine Learning and Deep Learning Techniques with Python. Apress, 2023.

[5] C. Quiroz-Vázquez, “Anomaly detection in machine learning: Finding outliers for optimization of business functions,” IBM Blog, https://www.ibm.com/blog/anomaly-detection-machine-learning/ (accessed Jun. 5, 2024).

[6] F. Lazzeri, Machine Learning for Time Series Forecasting with Python, 1st ed. Indianapolis: Wiley, 2021.

[7] J. Brownlee, “A gentle introduction to exponential smoothing for time series forecasting in Python,” http://MachineLearningMastery.com , https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/ (accessed Jun. 4, 2024).