**ENS 492 – Graduation Project (Implementation)**

**Progress Report II**

**Project Title: Prediction of Remaining Life of the Aircraft Engine**

**Group Members: Berfu Aydın – 23520**

**Ahmet Rauf Yavuz - 23887**

**Supervisor(s): Ahmet Demirelli**

**Date: 10/04/2021**



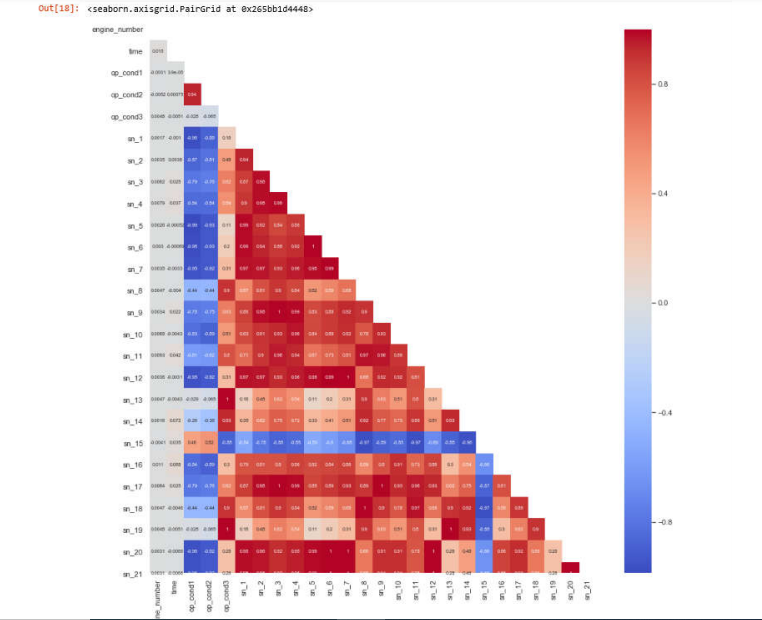
1. **PROJECT SUMMARY**

Aviation sector is getting more and more important today. In addition to building aircraft, their maintenance is quite difficult and costly. Aircraft that are not maintained or whose condition is uncertain cause irreparable casualties. Our starting point in this project: It is to prevent high costs and loss of life by ensuring that the aircraft that are close to the end of their engine life are detected in advance and the engine supply is made as quickly as possible.

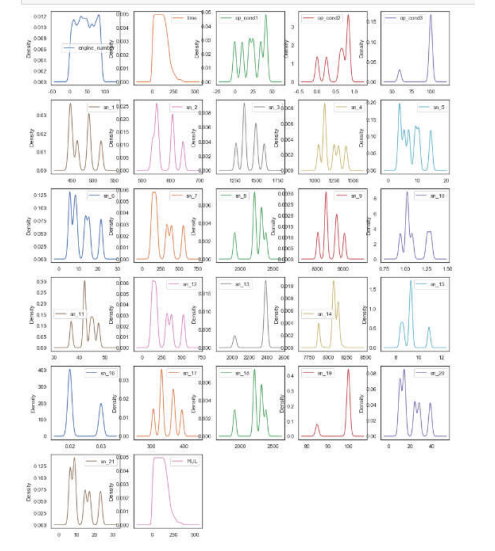
Our project aims to create machine learning algorithms to predict the remaining life of the aircraft engine based on approximately 20 sensor data obtained from jet aircraft engines. While doing this, it will be determined which model will work more effectively and with high accuracy by using different RNN models available in the literature. In this process, synthetic data we have obtained from NASA will be used.

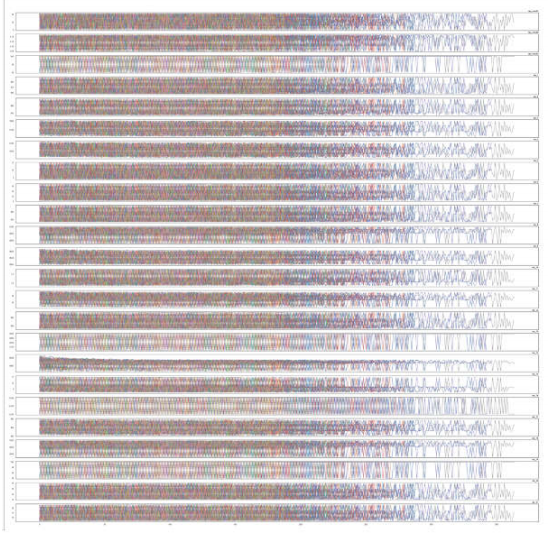
1. **SCIENTIFIC/TECHNICAL DEVELOPMENTS**

The aim so far is to recognize the available data and to find correlations between them by visualizing these data. This part, called data processing, is one of the most important parts before the machine learning model is created. In this way, the model is purified from unnecessary data and creates a measure for the result to be more consistent and for possible problems to be encountered in the future.



Heatmap is formed which is a very useful visualization tool, generally shows the correlation between  
features. The red colored boxes indicate that the correlation is intense. It can be seen that there are  
many features that have positive correlation in general. Light colored (eg white colored) boxes represent low correlation. For example, the value between sn\_9 and sn\_3 sensors is 1. This represents a complete  
positive correlation between these sensors. Which features will be used will be determined by considering all these correlations. Unnecessary features will not be included in the model.





This chart is one of the most efficient. Some data contains too much noisy data. This data does not need to be included in the model because it both deflects the result and creates unnecessary load for the model.

When looking at the general distribution of the feature values, it can be seen that the change depending on the specific motors is very intense in each feature. It seems difficult to find a general trend. However, it seems a little more likely that it can be found a specific trend for sn\_8, sn\_11, sn\_14, sn\_15 and sn\_18 sensors.

While creating the RNN (Recurrent Neural Network) model, different machine learning algorithms available in the market will be used. In this way, it will be observed which model gives better results.

1. **ENCOUNTERED PROBLEMS**

The two most likely problems to be encountered are under fitting and overfitting problems. These are the problems that are generally encountered and cause the result to be inefficient.

In short, the overfitting problem is that the model sticks to existing data too much. Another saying is that it starts to memorize rather than learning. In other words, the algorithm moves away from solving the main problem and focuses on attaching the values given to it. it focuses so much that it fits the values perfectly, but therefore it avoids solving the problem it has to solve.

Let me try to explain this with a concrete example from daily life. Imagine a student going to take the exam. But this student memorized the answers of previous exam questions, rather than learning the insight of topics a few days before the exam. When we put new questions in front of this student, this student will answer incorrectly because s/he memorizes the past questions. Likewise, if the model we created is very sticked to data in the training set, it will give wrong results when given the test data.

Under fitting, on the other hand, means that if a model has under-learning, as opposed to over-learning, the model does not fit the training data and therefore misses trends in the data.

There are some ways to eliminate the overfitting problem.

1. **Reducing the network’s capacity**

Our first model has a large number of trainable parameters. As this number  increases, the model memorizes the target class so easily for each training  example. As a result, this is not ideal for generalizing on new data. By lowering the capacity of the network, you force it to learn the patterns that  matter or that minimize the loss.

1. **Applying regularization**

Regularization techniques are generally used to increase performance by  preventing overfitting in the designed model. In addition, there are cases where it is used to reduce the complexity of the model without degrading performance.

1. **Train with more data**

Even, it won’t work every time, but training with more data can help algorithms detect the signal better.

1. **Remove Features**

Our model can be simplified by removing unnecessary features. For this, algorithms such as random forest can be used.

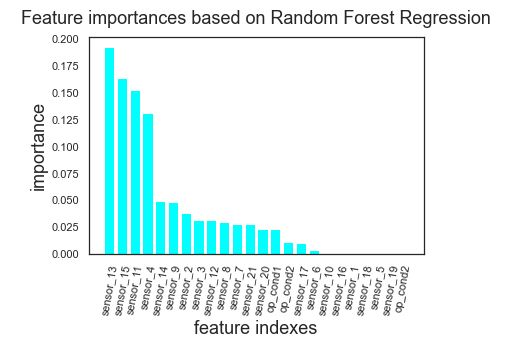
1. **Early Stopping**

In order to prevent unnecessary over-learning, the algorithm stops learning as soon as the model starts to memorize. This prevents too much reliance on data.

1. **SOME COMPLETED WORKS AND OUR FINDINGS**

There is some completed works that we can access and see their results for comparing with our work.

One of them is a Project that held on 2008 with PHM data, they got an 84.19% of accuracy. There is another project that uses stacked sparse Auto encoder. They obtained 83.82% accuracy which is close to the first project that we mentioned. One of the best projects that related to our project has quite high accuracy which is 95% regression accuracy. They used PSO-SVM based algorithm in this project4.



We have determined which feature is more important using Random Forest, which is usually applied here. As can be seen in the figure, sensor 13 data is the most important feature of our data set.

1. **TASKS TO BE COMPLETED UNTIL FINAL REPORT**

Until Final report, roughly each week, we will try to build different ML models to compare each one to another for deciding which one is more accurate for this type of task and dataset. So far, we tried Random Forest algorithm. We will try to implement other ML models and compare them among each other.

1. **REFERENCES**

1.<https://medium.com/@gulcanogundur/overfitting-a%C5%9F%C4%B1r%C4%B1-%C3%B6%C4%9Frenme-underfitting-eksik-%C3%B6%C4%9Frenme-ve-bias-variance-%C3%A7eli%C5%9Fkisi-b92bef2f770d>

2.

<https://medium.com/@gulcanogundur/overfitting-a%C5%9F%C4%B1r%C4%B1-%C3%B6%C4%9Frenme-underfitting-eksik-%C3%B6%C4%9Frenme-ve-bias-variance-%C3%A7eli%C5%9Fkisi-b92bef2f770d#:~:text=A%C5%9F%C4%B1r%C4%B1%20%C3%B6%C4%9Frenmenin%20aksine%2C%20bir%20model,verilerdeki%20trendleri%20ka%C3%A7%C4%B1rd%C4%B1%C4%9F%C4%B1%20anlam%C4%B1na%20gelir.&text=Underfitting%20sorunu%20olan%20modellerde%20hem,ve%20y%C3%BCksek%20bias'a%20sahiptir>.

3.

<https://www.hindawi.com/journals/complexity/2018/3813029/#copyright>

4. Nieto, PJ García, et al. "Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability." *Reliability Engineering & System Safety* 138 (2015): 219-231.