Those responsible for the project are very excited about the initiative and, when asking for a technical proof of concept, they have put forth as main requirements:

* Can we reduce our expenses with this type of maintenance using AI techniques?

Sim, é possível reduzir os gastos com manutenção do sistema de ar usando machine learning (ML). Utilizando modelos de ML é possível observar quais variáveis possuem maior efeito no sistema de ar dos caminhões. Com essas informações podemos selecionar essas variáveis e colocá-las na categoria de que se houver alterações em seus valores deverão ser levados diretamente para manutenção para que não haja custo de $500 para correção do problema. Podemos também selecionar as variáveis que menos afetam o sistema de ar dos caminhões. Assim, podemos colocar essas variáveis na categoria de que se houver alterações em seus valores, não há necessidade de ir imediatamente para manutenção, não havendo o custo de $10. Resumindo, podemos prever quais caminhões têm maior probabilidade de apresentar falhas no sistema de ar. Isso permitirá que a manutenção preventiva seja realizada de forma mais eficiente, evitando custos elevados de manutenção corretiva.

* Can you present to me the main factors that point to a possible failure in this system?

O principal fator que causa a possibilidade de problema no sistema de ar é a variável “aa\_000” que foi apontada como característica mais importante para alteração no sistema de ar em todos os modelos. Outras variáveis mencionadas para cada componente gerado pela PCA (Análise de Componentes Principais):

Variáveis mais importantes para componente (PC1): bx\_000, cc\_000, by\_000, ba\_003, cs\_005, ba\_004, ci\_000, ba\_002, ag\_005 e dc\_000.

Variáveis mais importantes para componente (PC45): aa\_000, dj\_000, dk\_000. ar\_000, cy\_000, ay\_001, cs\_001, az\_009, dy\_000 e ab\_000.

Variáveis mais importantes para componente (PC2): ag\_003, ag\_004, cn\_002, cn\_001, ay\_006, am\_0, al\_000, ag\_006, ag\_002 e ag\_007.

**Challenge Activities**

To solve this problem we want you to answer the following questions:

1. What steps would you take to solve this problem? Please describe as completely and clearly as possible all the steps that you see as essential for solving the problem.

Para resolver esse problema eu primeiro absorveria o máximo de informações sobre as bases de dados, depois eu faria uma análise exploratória dos dados. Seria necessário tratar valores nulos, converter variáveis, aplicar técnicas de amostragem, e normalizar a base de dados. Depois eu separaria os dados em conjunto de treino e teste para treinar os modelos de machine learning e usaria métricas para avaliar a performance dos modelos usados. Usaria técnicas para selecionar variáveis que possuem maior impacto na variável “class”. E por último implementaria o modelo e monitoraria sua performance.

1. Which **technical** data science metric would you use to solve this challenge? Ex: absolute error, rmse, etc.

AUC-ROC, precisão, recall e F1-score.

1. Which business metric ***would*** you use to solve the challenge?

Redução dos custos totais de manutenção do sistema de ar.

1. How do technical metrics relate to the business metrics?

A performance dos modelos preditivos, que é medida pelas métricas técnicas, está diretamente relacionada a capacidade de prever falhas corretamente, o que influencia a redução de custos de manutenção ao evitar manutenções corretivas caras.

1. What types of analyzes would you like to perform on the customer database?

Análise de correlação, análise de distribuição de variáveis, análise de outliers e análise de impacto de variáveis nos custos de manutenção.

1. What techniques would you use to reduce the dimensionality of the problem?

PCA (Análise de Componentes Principais).

1. What techniques would you use to select variables for your predictive model?

PCA (Análise de Componentes Principais) e Feature Importances

1. What predictive models would you use or test for this problem? Please indicate at least 3.

Regressão logística, Random Forest e XGBoost

1. How would you rate which of the trained models is the best?

Comparando as métricas de AUC-ROC, precisão, recall e F1-score nos conjuntos de validação e teste.

1. How would you explain the result of your model? Is it possible to know which variables are most important?

Usando técnicas de explicabilidade como SHAP, Permutation Feature Importance e Feature Importances para mostrar a contribuição de cada variável na predição.

1. How would you assess the financial impact of the proposed model?

Calculando a redução dos custos de manutenção ao evitar manutenções corretivas e otimizar manutenções preventivas.

1. What techniques would you use to perform the hyperparameter optimization of the chosen model?

Grid Search e Random Search.

1. What risks or precautions would you present to the customer before putting this model into production?

Verificar a qualidade dos dados, monitorar continuamente a performance do modelo e ajustar conforme necessário, e considerar as implicações de possíveis vieses no modelo.

1. If your predictive model is approved, how would you put it into production?

Utilizando ferramentas como Docker para containerização e APIs para integração com o sistema existente.

1. If the model is in production, how would you monitor it?

Monitorando métricas de performance e validando as predições do modelo periodicamente.

1. If the model is in production, how would you know when to retrain it?

Monitorando a queda na acurácia das predições e mudanças nos dados de entrada que indicariam a necessidade de atualização do modelo.