Análise de Eficiência Energética e Uso Racional de Recursos pelos Equipamentos de Ar Condicionado Manutenidos pelo CEMEQ UNICAMP.

Palavras-Chave: Eficiência Energética, *Machine Learning*, Ciência de Dados.

Autores:

Gleyson Roberto do Nascimento FEEC/UNICAMP.

Prof. Dr. Luiz Carlos Pereira da Silva (Orientador) FEEC/UNICAMP.

Prof. Dr. Hildo Guillardi Júnior (Coorientador) UNESP/SJBV.

INTRODUÇÃO:

De junho de 2013 até dezembro de 2021 foram registradas quarenta mil e duzentas ordens de serviço de manutenção no Banco de Dados de Ordens de Serviço do CEMEQ UNICAMP e, pela Figura 1, destas, 39% eram referentes aos Equipamentos de Condicionamento de Ar e Climatização de Ambiente (38% manutenção e 1% compra) e estes são responsáveis por grande parte da parcela da Eficiência Energética e Uso Racional de Recursos pela Universidade. Com esta alta demanda surge o questionamento: até que ponto é viável a manutenção destes equipamentos do ponto de vista de Eficiência Energética e Uso Racional de Recursos?

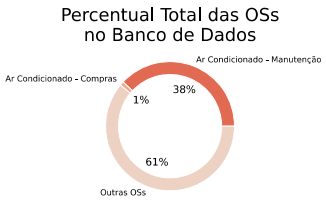

Figura 1 – Percentual dos tipos de Ordens de Serviço no Banco de Dados do CEMEQ UNICAMP.

Na vanguarda deste e de outros questionamentos, surge em 2017 o Projeto Campus Sustentável, que desde então vem proporcionando à UNICAMP uma grande melhora no uso de recursos, sustentabilidade e credenciando a instituição junto às mais modernas práticas de *Smart Campus* e *Smart Cities*, assim, considerando este contexto, este projeto se propõe a realizar um estudo mais aprofundado do patrimônio de Equipamentos de Ar Condicionado e Climatização manutenidos pelo CEMEQ UNICAMP com relação ao impacto deste patrimônio no escopo geral da Universidade.

Desta forma, sob a ótica do Programa Campus Sustentável, foram avaliadas a Eficiência Energética e o Uso de Recursos por parte deste patrimônio registrado e, foi proposto um modelo computacional para uma melhor tomada de decisão quando um dado Equipamento de Ar Condicionado e Climatização de Ambiente necessitar de manutenção no cenário da UNICAMP. Tal estudo ocorreu através de processos de *Data Science* e *Machine Learning*.

METODOLOGIA:

Inicialmente, para que este projeto fosse viável, além do Banco de Dados do CEMEQ UNICAMP, foi necessário o acesso a outros bancos de dados da UNICAMP para a complementação de informações e melhor modelamento do banco final de trabalho. Além disso, como não há padronização na Universidade, foi muito necessária a colaboração de funcionários da UNICAMP para que as variáveis e estruturas de banco pudessem ser compreendidas.

Uma vez que as condições iniciais foram reunidas, foi utilizada a metodologia *Cross-industry standard process for data mining* (CRISP-DM), que atualmente é uma das mais difundidas técnicas em *Data Science*, de forma que, além da Revisão Bibliográfica, foram realizadas sete etapas técnicas:

1) *Business Understanding*: Compreensão da totalidade do problema para a busca da melhor abordagem para responder às perguntas de pesquisa;

2) *Data Understanding*: Avaliação do Banco de Dados de Ordens de Serviço de Equipamentos de Ar Condicionado e

Climatização Manutenidos pelo CEMEQ UNICAMP e busca de Datasets e variáveis que contribuam para responder às perguntas de pesquisa;

3) *Data Preparation*: Limpeza, uniformização, normalização e adequação da base de dados final de trabalho;

4) *Data Mining:* Através de técnicas estatísticas como testes de correlação, p-valor, análise gráfica (*scatter plot*, histograma, *q-q plot*), realizar a "mineração de dados" em busca do comportamento das variáveis e de dados que possam vir a contribuir para a estrutura do projeto;

5) *Modelling:* Através de transformação de dados, como redução de ordem com *Principal Component Analysis* (PCA), clusterização e regressão linear por *Machine Learning*, buscar um modelo computacional para compreender a correlação e peso entre as variáveis, consequentemente, os critérios que determinam a viabilidade de manutenção dos equipamentos no cenário da UNICAMP;

6) *Evaluation:* Através de técnicas de validação cruzada em *Machine Learning*, buscar validar o modelo computacional encontrado;

7) *Deployment:* Avaliação das melhores metodologias para implementação do uso do modelo computacional no Campus da UNICAMP e receptividade pelos profissionais da Universidade.

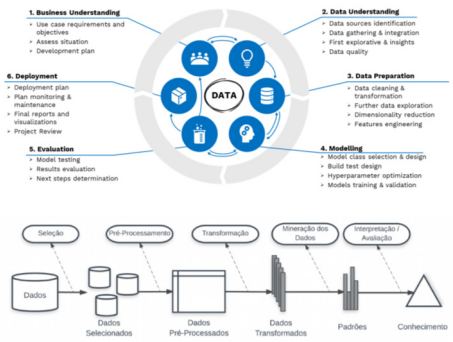

Figura 2 – Fluxograma Geral da Metodologia CRISP-DM para Análise de Dados.

RESULTADOS E DISCUSSÃO:

a) *Revisão Bibliográfica*: Durante o processo de Revisão Bibliográfica, foram analisadas mais a fundo as referências iniciais utilizadas na Proposta de Trabalho e, também, através das bibliotecas da Capes, IEEE, Mendeley e Google Acadêmico foram incorporadas novas referências, de forma a compor a Bibliografia final;

b) *Business Understanding, Data Understanding e Data Mining*: Este projeto, de forma mais simplificada, pode ser entendido como um classificador por *Machine/Deep Learning* com a análise da eficiência energética destes equipamentos, funcionando como um peso de tomada de decisão, assim, para que fosse possível a predição e consequentemente responder as perguntas de pesquisa deste projeto, foram necessários as seguintes bases de dados: banco de manutenção dos equipamentos no CEMEQ UNICAMP; banco de patrimônio de equipamentos da UNICAMP; banco de dados do Selo Procel, banco da potência dos equipamento; banco das contas de consumo de energia da UNICAMP; banco do custo de aquisição de equipamentos para a UNICAMP.

Após a junção dos bancos de dados, chegou-se a um *dataset* com as seguintes correlações e p-valores:

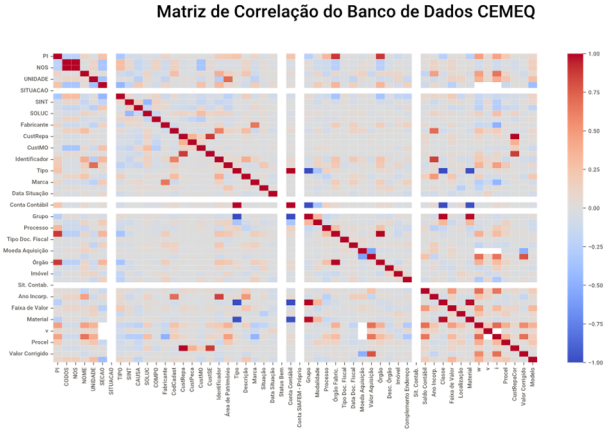

Figura 3 – Matriz de correlações entre as variáveis.

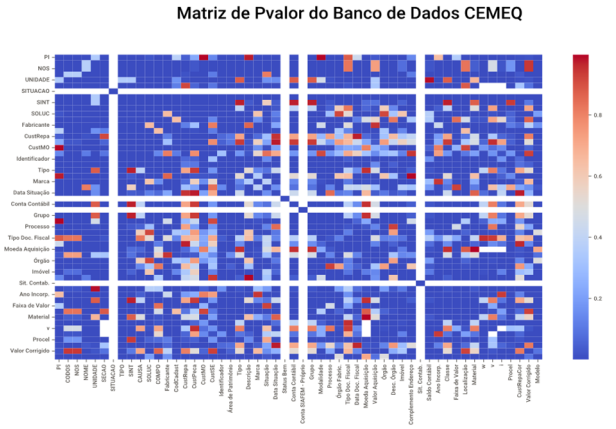

Figura 4 – Matriz de p-valor entre as variáveis.

Com os resultados de correlações e p-valor, foi possível seguir os procedimentos de *Data Mining* citados na seção de

metodologia, resultando num *dataset* com 4449 linhas e 52 *features* e 1 *target* com 4 categorias distintas, sendo elas: 1) troca em 2022; 2) troca em 2025; 3) manutenção em até 5 anos; 4) manutenção em até 10 anos. Tais categorias foram escolhidas considerando os seguintes dados:

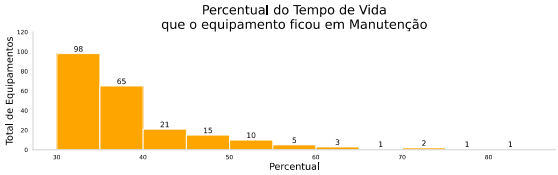

Figura 5 – Equipamentos que ficaram mais que 30% do tempo de vida em manutenção.

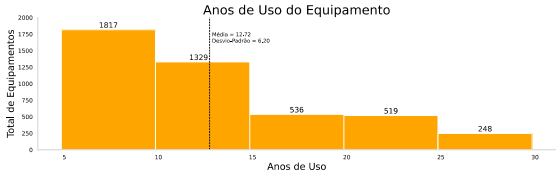

Figura 6 – Histograma dos anos de uso dos equipamentos.

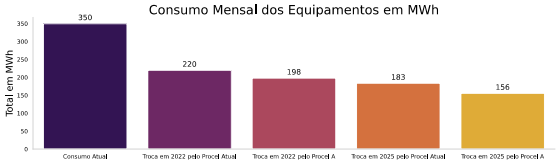

Figura 7 – Cenários considerando o atual consumo energético (MWh) e considerando a troca de equipamentos.

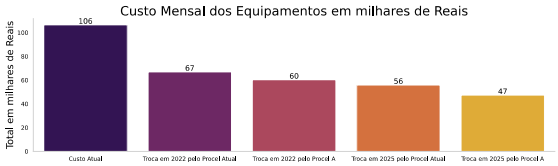

Figura 8 – Cenários considerando o atual custo financeiro (milhares de Reais) e considerando a troca de equipamentos.

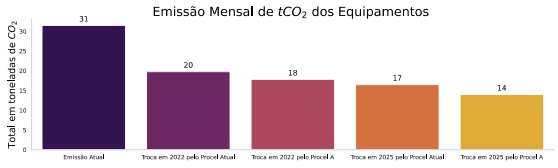

Figura 9 – Cenários considerando a atual emissão de carbono (toneladas de CO_2) e considerando a troca de equipamentos.

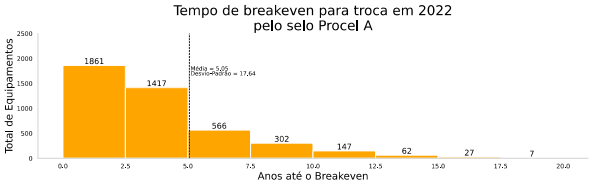

Figura 10 – Tempo de breakeven considerando a troca de equipamentos em 2022 por equipamentos selo Procel A.

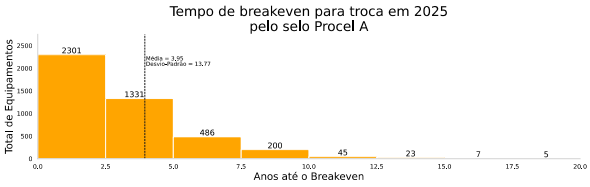

Figura 11 – Tempo de breakeven considerando a troca de equipamentos em 2025 por equipamentos selo Procel A.

Desta forma, considerando que há uma quantidade relevante de equipamentos que permaneceu mais que 30% do tempo de vida em manutenção; que a maioria dos equipamentos que receberam manutenção possuem mais de 10 anos de uso e, que pela Instrução Normativa SRF nº 162 de 31/12/1998, a Receita Federal estabelece que a depreciação para este tipo de bem é de 10% a.a.; que a troca por equipamentos de selo Procel A levaria a uma economia de 45% no consumo energético, emissão de carbono e custo financeiro e que essa troca, em média, teria um *breakeven* entre 4 a 6 anos, optou-se pelas 4 categorias mencionadas;

c) *Modelling, Evaluation e Deployment*: com apenas 4449 linhas, não era viável manter um dataset com 52 *features* e 1 *target*, dada a probabilidade de *overfitting*, desta forma, para os procedimentos em *Machine Learning* optou-se por duas abordagens distintas: a primeira fazendo apenas a *feature selection* utilizando o algoritmo *Boruta*, o que levou a 33 *features* e 1 *target*, já a segunda abordagem fazendo uso de PCA com 15 componentes e 4 *clusters*.

Considerando a classificação não binária, optou-se pelo uso das seguintes opções de classificação: OvO com SVC, OvR com SVC, KNC, *Random Forest*, *Extra Tree*, *Decision Tree*, *Gradient Boost* e MLP. O *dataset* foi dividido em 70% treino, 30% teste, houve otimização de hiperparâmetros por *Randomized Grid Search* e houve *pipeline* de normalização e validação cruzada com *KFold* com 10 *folds*.

Após o treinamento, os modelo com melhor desempenho foi o MLP com aplicação do *Boruta* e sem PCA com os seguintes hiperparâmtros: *early stopping* = *True*, *random state* = *np.random(123)*, *hidden layer sizes* = (4,), *solver* = *'lbfgs'*, *learning rate* = *'invscaling'*, *learning rate init* = 0.01, de forma que o modelo foi avaliado com os seguintes parâmetros de *report*:

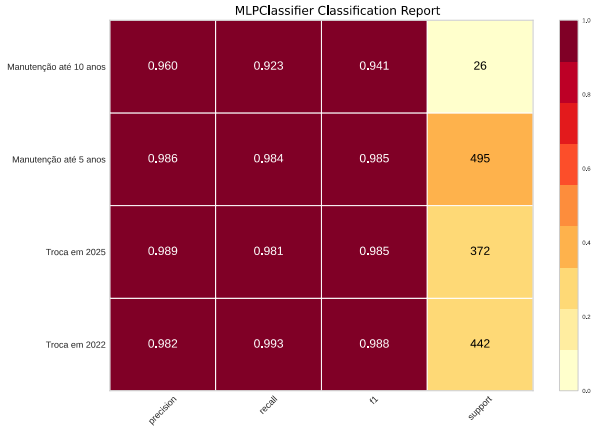

Figura 12 – Matriz de report do MLP Classifier.

Além disso, para este modelo, no *cross validation* foram obtidos: *score* médio: 0.9833, *score* mínimo: 0.95833, *score* máximo: 0.99356, desvio padrão: 0.98397 com a seguinte matriz de confusão:

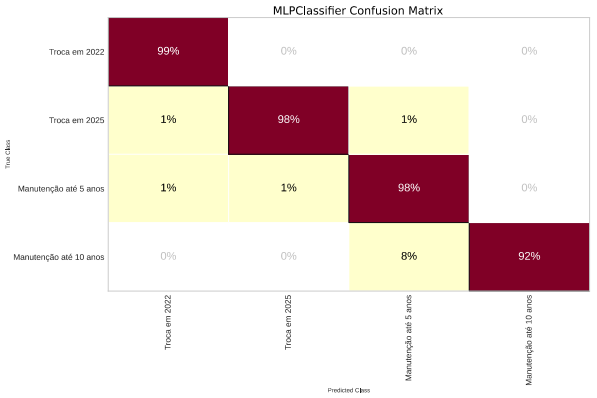

Figura 13 – Matriz de confusão do MLP Classifier.

Por fim, considerando o cenário de integração das bases de dados necessárias para que este modelo de *MLP Classifer* possa ser colocado em *pipeline* de produção, não foi viável a construção de *software* em modelo *MLOPS*, sendo apenas possível o uso do modelo na construção de *dashboards* com informações acerca dos modelos e a sua respectiva classificação nas 4 classes disponíveis.

CONCLUSÕES:

Através deste projeto foi possível fazer uma análise do cenário existente para os equipamentos de ar condicionado que são manutenidos pelo CEMEQ UNICAMP, assim, verificou-se as necessidades de melhorias gerais com relação aos processos e equipamentos.

Com relação aos processos, notou-se que a Universidade, grande referência em termos acadêmicos, ainda não possui estrutura de fácil acesso e integração aos dados públicos e de interesse de pesquisa, sendo necessária uma atuação significativa do orientador e coorientador no convencimento dos departamentos e nas tratativas burocráticas envolvidas.

Como muitos dos bancos de dados não são estruturados e possuem a possibilidade de erros sistemáticos na entrada de dados, não podemos afirmar de forma categórica que a confiança nos dados coletados é absoluta, sendo bastante improvável que equipamentos tenham ficado 80% do seu tempo de vida em manutenção, por exemplo.

Assim, é aconselhável que as bases de dados passem a estruturas formais de bancos de dados relacionais, com integração entre as unidades da UNICAMP, com formalismo, dicionário e *dtypes* centralizados e organizados por departamento técnico adequado ao projeto.

Sobre os equipamentos, a primeira medida a ser adotada é o controle do tempo máximo de vida útil de um equipamento, de forma que, considerando a Instrução Normativa da Receita Federal que estabelece a depreciação deste tipo específico de bem em 10% a.a., fazendo com que equipamentos com mais de 10 anos já tenham seu custo totalmente depreciado, fato que leva a uma defasagem tecnológica e, consequentemente, a ineficiência energética e diminui o uso racional dos recursos públicos da Universidade.

Além disso, foi possível verificar que, ao contrário do que é convencional, os equipamentos com maiores índices de pedidos de manutenção não são os mais antigos, o que pode sugerir um problema pontual com marcas/modelos ou nas licitações, demonstrando uma necessidade de maior controle de qualidade na aquisição e recebimento dos mesmos.

Outra medida a ser adotada é a troca de equipamentos que possuem mais de 20 anos de uso ou que permaneceram por mais de 30% do seu tempo de vida em manutenção, considerando a questão de depreciação, faltas de peças de manutenção no mercado, ineficiência energética e elevada taxa de ausência no departamento para o qual o equipamento é necessário.

Considerando o cenário atual e a mudança de eficiência dos equipamentos com selo Procel a partir de dezembro/2022 e dezembro/2025, foi possível verificar que na troca dos equipamentos com os perfis específicos, é possível economizar aproximadamente 45% em termos financeiros, de consumo energético em MWh e em emissão de toneladas de carbono, com tempo médio de *breakeven* da compra dos equipamentos entre 4 a 6 anos.

Já com relação ao aprendizado técnico científico, o que ficou evidenciado é que, para estes dados, com um *target* do tipo *multiclass*, das opções recomendadas e implementadas, a que possuiu

melhor desempenho foi o *MLP Classifier*, de forma que mesmo uma grande quantidade de tempo gasto na otimização de hiperparâmetros dos modelos de *Machine Learning* por pipeline com *Randomized Grid Search* não conseguiu ter melhor performance que o *MLP Classifier* com 4 camadas ocultas, sendo interessante o uso do *Deep Learning* para a solução deste problema de classificação.

Além disso, no caso específico das metodologias de *Machine Learning*, para estes dados, a *feature selection* com o algoritmo *Boruta* trouxe melhores resultados que o PCA em 15 componentes, fato que nem sempre se evidencia.

BIBLIOGRAFIA:

[1] Assessoria de Economia e Planejamento (AEPLAN). **Proposta de Distribuição Orçamentária 2022**. < https://www.aeplan.unicamp.br/proposta_orcamentaria/unicamp/pdo_unicamp_2022.pdf, acesso em abril de 2022>;

[2] Larran Jorge, M.L., Herrera Madueno, J., Calzado Cejas, M.Y., Andrades Peña, F.J.; **An approach to the implementation of sustainability practices in Spanish universities**, J. Clean. Prod. 106, 34e44. 2015;

[3] Schmitt-Figueiro, P., Raufflet, E., **Sustainability in Higher Education: a systematic review with focus on management education**, J. Clean. Prod. 106, 22e33. 2015;

[4] Veli-Pekka N., Seija K., Mats E., Karl A.H.; **State-of-the-art and good practice in the field of living labs**, 2006;

[5] Yang, M.; Nazir, S.; Qingshan, X; Shaukat, A.; **Deep learning algorithms and multicriteria decision-making used in big data: a systematic literature review, Wiley-Hindawi Complexity**, Vol. 2020, 2020;

[6] Khorasgani, H.; Wang, H.; Gupta, C.; Farahat, A.; **An Offline Deep Reinforcement Learning for Maintenance Decision-Making,** Annual Conference of the Prognostics and Health Management Society, 2021;

[7] Bolsa Eletrônica de Compras – SP, **Pregão Eletrônico** < https://www.bec.sp.gov.br/bec_pregao_UI/OC/pregao_oc_pesquisa.aspx?chave= , acesso em setembro de 2021>;

[8] Knaflic, C. N.; **Storytelling with Data: A Data Visualization Guide for Business Professionals**, 1º Ed., Wiley;

[9] Geron, A.; **Hands–On Machine Learning with Scikit–Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**, 2º Ed., O'Reilly.

[10] Harris, C.R., Millman, K.J., van der Walt, S.J. **Array programming with NumPy**. Nature 585, 357–362 (2020). DOI: 0.1038/s41586-020-2649-2.

[11] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, ˙I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. **SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods**, vol. 17, pp. 261–272, 2020.

[12] T. pandas development team, **"pandas-dev/pandas: Pandas**," Feb. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.3509134

[13] J. D. Hunter. **Matplotlib: A 2d graphics environment**, Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007.

[14] M. Waskom and the seaborn development team, "mwaskom/seaborn," Sep. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.592845

[15] Gulli A, Pal S. **Deep learning with Keras**. Packt Publishing Ltd; 2017.

[16] Pedregosa, F. **Scikit-learn: Machine Learning in Python**. JMLR 12, pp. 2825-2830, 2011.

[17] Abadi, M., Barham P, Chen, J., Chen, Z., Davis, A., Dean, J. **Tensorflow: A system for large-scale machine learning**. In: 12th Symposium on Operating Systems Design and Implementation . 2016. p. 265–83.

[18] Imprensa Nacional, **Portaria nº 234: Aperfeiçoamento parcial dos Requisitos de Avaliação da Conformidade para Condicionadores de Ar** <https://www.in.gov.br/web/dou/-/portaria-n-234-de-29-de-junho-de-2020-264423659, acesso em setembro de 2021>.