|  |  |  |  |
| --- | --- | --- | --- |
| Projeto: Soft controle de gramatura | | PDSA nº: 2 | Data: 12/01/2024 |
| Objetivo:  - Garantir que os dados estejam preparados para análise e manipulação avançada, permitindo que técnicas mais sofisticadas, como Regressão Linear Múltipla ou Random Forest, sejam aplicadas de maneira eficiente. | | | |
| PLAN | | | |
| Questões | Predições | | |
| 1. Identificar e tratar outliers que possam distorcer o modelo. | Os dados que forem identificados serão descartados para preservar a linearidade e evitar distorções para o modelo preditivo. | | |
| 2.Normalizar variáveis para modelos que exigem escalas semelhantes | Serão executadas conforme a necessidade e adaptação dos dados ao contexto do objetivo de prever a gramatura. | | |
| 3.Verificar a correlação entre variáveis para eliminar redundâncias ou variáveis menos informativas. | Serão utilizados métodos analíticos estatísticos e utilização de algoritmos. | | |
| 4.Criar e validar variáveis derivadas que possam melhorar o desempenho do modelo. | Será feito conforme a necessidade ou uso de algoritmos para modelo. | | |
| 5. Implementar validação cruzada para garantir estabilidade na previsão. | Será usada a validação padrão | | |
| 1- Outliers serão tratados com Z-Score que verifica desvios padrão extremos em relação à média.  2- As variáveis serão dimensionada na mesma escala para otimizar o aprendizado. A normalização ou padronização garante que nenhuma variável tenha influência desproporcional através da técninca de MinMaxScaler ou Normaliza os valores entre 0 e 1, preservando proporções.  3- As correlações será analisadas através de gráficos de Heatmap, hexbin  4- Variáveis derivadas podem capturar relações complexas que as variáveis originais não conseguem evidenciar.  5 -A validação cruzada avalia o desempenho do modelo em diferentes subconjuntos de dados, garantindo previsões mais estáveis e generalizáveis. Será utilizada a K-Fold Cross-Validation que faz divisão dos dados em k partes, testando e treinando em combinações diferentes. | | | |
| **DO the Action Plan** | | | |
| 1- fazer a remoção de outliers com metodo drop() passando os indices pela busca atraves de strs, como ‘Bad’, ‘System out’,  2-fazer a normalização com zscore e min max ,  3-Aplicar o mapa de com metodo .corr do pandas e aplicar o heatmap e hexbin.  4-Não fazer testes com taxas e variáveis derivadas.  5- Utilizar o Kfold para aplicar a validação cruzada | | | |
| **STUDY** | | | |
|  | | | |
| 1. Melhorar a interação com sistemas que extraem dados com .xml. | | | |
| 2. Melhorar o uso de outras ferramentas de estatisticas como ANOVA e etc. | | | |
| 3. | | | |
|  | | | |
| **ACT** | | | |
| Observar mais padrões e possibilidades de ajuste com variáveis derivadas. | | | |
| Qual será o objetivo do próximo ciclo PDSA?  Avaliar outros algoritmos que possibilitam opções de melhoria do modelo. | | | |