

Aplicação de Ciência de Dados em informações sobre gestão de pessoas do Poder Executivo Federal disponibilizadas pelo Ministério da Economia.

Antonio Alef Oliveira Farias¹, Aldemir Melo Rocha Filho¹, Jonas Amancio Silva Santos¹, Tiago Paulino Santos¹

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Maceió – AL – Brazil

aaof@ic.ufal.br, amrf@ic.ufal.br, jass@ic.ufal.br, tps2@ic.ufal.br

Abstract. *This article aims to analyze information on people management of the Federal Executive Branch, and through machine learning techniques, result in a prediction of vacancy in public agencies. Resulting in useful information for the educational market focused on contests, and citizens interested in the area. In this way, all the steps of data science were carried out, from cleaning the database to the application of a learning model based on SVM (Support Vector Machine).*

Resumo. *Este artigo tem como objetivo analisar informações sobre gestão de pessoas do Poder Executivo Federal, e através das técnicas de machine learning, resultar em uma previsão de vacância em órgãos públicos. Resultando em informações úteis para o mercado educacional focado em concursos, e cidadãos interessados na área. Desse modo, foram feitos todos os passos de ciência de dados, da limpeza de base dados à aplicação de um modelo de aprendizagem baseado em SVM (Máquina de Vetor de Suporte - Support Vector Machine).*

1. Aplicação

Diante dos dados fornecidos pelo ministério da economia, o presente trabalho tem como proposta extrair informações a fim de obter os órgãos governamentais e regiões onde possivelmente haverá uma vacância de cargos públicos. Desse modo, através das técnicas de machine learning obtém-se a estimativa de tempo de trabalho restante, até que os servidores se aposentem.

2. Experimentos

2.1. Base de dados

Os dados analisados disponibilizado pelo ministério da economia consistem em um conjunto de dados cadastrais de aposentados sob o regime jurídico único do Poder Executivo Federal Civil, como nome, portaria e tipo de aposentadoria. Além disso, há também os dados dos servidores públicos federais que recebem abono permanência, como nome, CPF, cargo, órgão.

Após a obtenção de todos os dados necessários relativos ao ano de 2020, iniciou-se a etapa de pré-processamento, ou seja, a utilização de técnicas para a limpeza, integração e redução desses dados. Foram feitas as seguintes mudanças nessa etapa:

- Após fazer o download de todos os CSVs relativos a todos os meses de 2020, combinou-se os dados de cada mês em um Data Frame. Essa etapa foi realizada tanto para os aposentados quanto para os abonos.
- Retirou-se as linhas com cabeçalho;
- Converteu-se a coluna de rendimento líquido do DataSet Aposentados;
- Separou-se os dados da coluna “Inicio_Abono” em outras duas novas colunas “Mês” e “Ano”. Após essa separação, retirou-se a coluna “Inicio_Abono”;
- Converteu-se os dados da coluna “Valor” de string para float;

2.2. Estatística descritiva e inferência

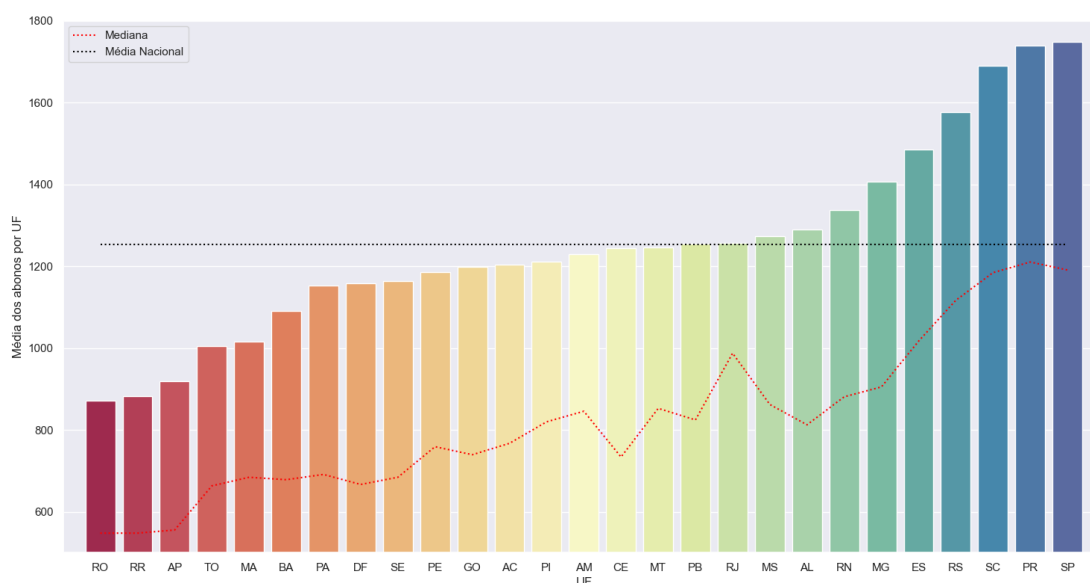


Gráfico 1: Valor médio mensal dos abonos por estado.

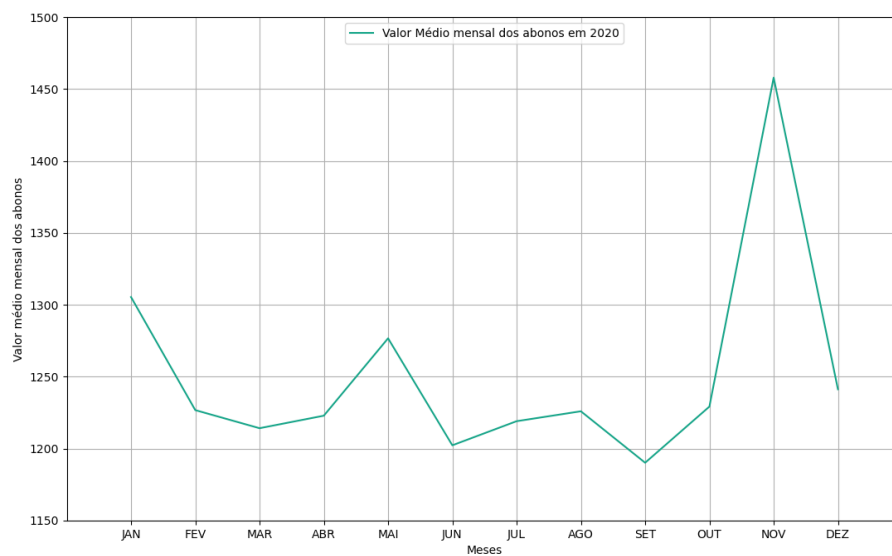


Gráfico 2: Valor médio mensal dos abonos em 2020.

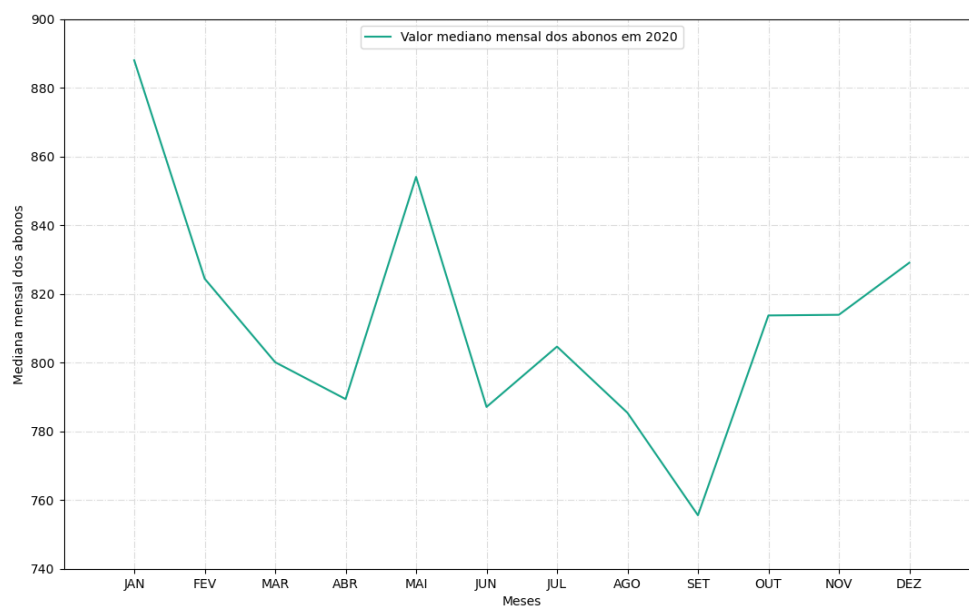


Gráfico 3: Valor mediano mensal dos abonos em 2020



Gráfico 4: Top 10 médias de rendimento por Órgão

2.3. Métodos e métricas avaliados

Em primeiro momento, precisou-se usar um encoder como método para discretizar as features de cargos e órgãos de maneira que se pudesse tratá-los dentro de um modelo de aprendizagem de máquina. Este procedimento foi executado nos dois dataframes (abonos e aposentados) para utilizar dentro da experiência proposta. Após criar o primeiro dataframe baseado nos dados dos aposentados foram criadas duas variáveis, a primeira contendo as features com os cargos e órgãos e a segunda contendo a variável alvo, o tempo de serviço público do aposentado.

Em razão do tamanho, para implementar o modelo melhorando a qualidade regressora utilizou-se o método de validação cruzada implementado através da biblioteca do SkLearning, o StratifiedKFold. Após alguns testes heurísticos para definir um valor para K para 5, obtendo assim um balanço entre custo computacional e qualidade na regressão.

Para realizar a regressão, foi utilizado um modelo de aprendizagem baseado em Máquina de Vetor de Suporte (SVM). Foram utilizadas também outros regressores baseados em redes neurais e Naive Bayes, contudo não obtendo uma boa generalização dos dados para o caso em questão, produzindo resultados bastantes ruidosos e claramente fora da realidade.

3. Resultados

Após a aplicação das técnicas, foi possível construirmos dois gráficos com o intuito de responder e prever situações futuras. O primeiro tenta prever o tempo estimado de trabalho restante até a aposentadoria, enquanto o segundo nos trás os órgãos com maior tendência de terem déficit em vagas num futuro não tão distante. Tais resultados levam em consideração uma análise baseada no cruzamento de informações, estas, presentes tanto nas tabelas de servidores em situação de abonos quanto na de servidores aposentados, possibilitando assim estimar as informações apresentadas a seguir:

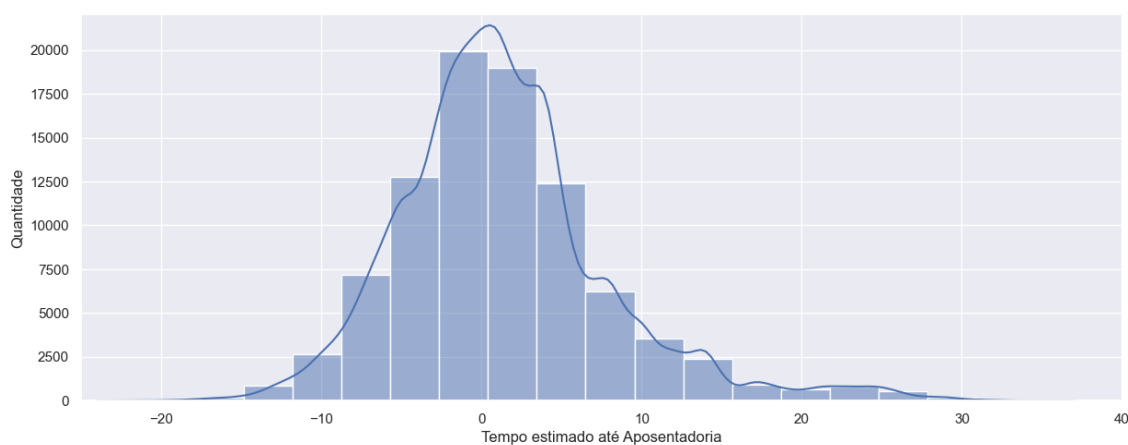


Gráfico 5: Histograma de tempo estimado até Aposentadoria

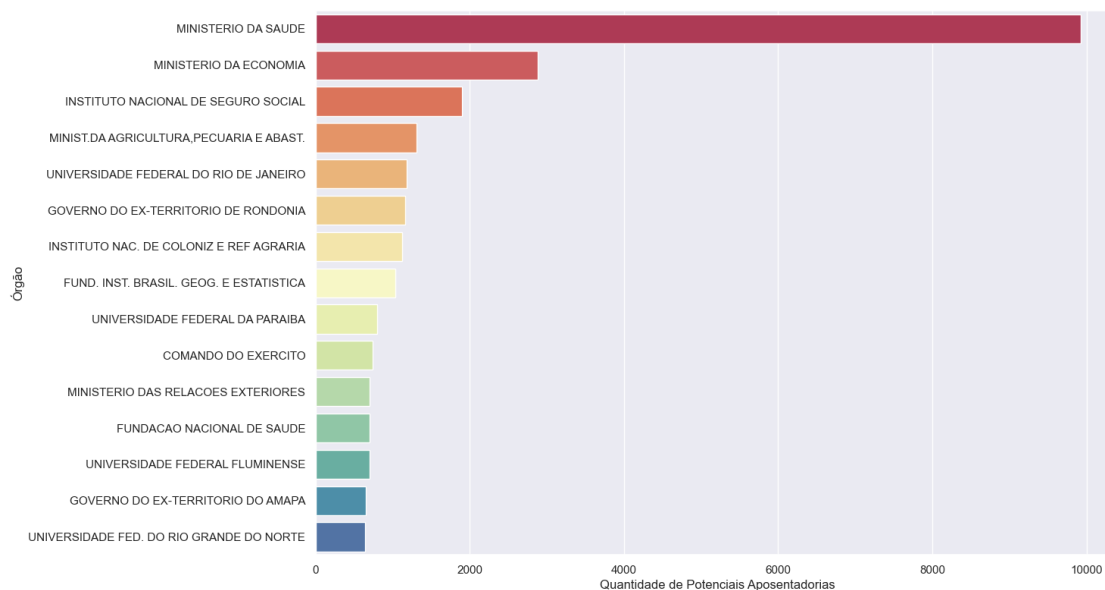


Gráfico 6: TOP 15 Órgãos de potencial vacância para 2021 e 2022

4. Conclusão

De acordo com os resultados obtidos no gráfico 06, percebe-se que o órgão com maior índice de vacância para os próximos anos é o ministério da saúde seguido do ministério da economia. Consequentemente, essa vacância pode resultar em uma queda na qualidade do serviço público do respectivo órgão governamental. Em contrapartida, essa futura falta de vagas gera a abertura de novos concursos públicos, fato que pode ser de grande interesse para um determinado público e para o mercado educacional focado em concursos.

5. Referências

Gestão de Pessoas (Executivo Federal) - Abono Permanência. Disponível em: <[Gestão de Pessoas \(Executivo Federal\) - Abono Permanência - Conjuntos de dados - Portal Brasileiro de Dados Abertos](#)>. Acesso em: 30 de agosto de 2021.

Gestão de Pessoas (Executivo Federal) - Aposentados. Disponível em: <[Gestão de Pessoas \(Executivo Federal\) - Abono Permanência - Conjuntos de dados - Portal Brasileiro de Dados Abertos](#)>. Acesso em: 30 de agosto de 2021.