

18

NÚMERO 1



REVISTA
**DIALOGO E
INTERAÇÃO**

ISSN 1275-3687



FACCREI



<https://www.faccrei.edu.br/revista>

MODELAGEM PREDITIVA DE IMPRESSÕES COM SKLEARN E GRADIENT BOOSTING: UMA ABORDAGEM PRÁTICA

PREDICTIVE PRINT MODELING WITH SKLEARN AND GRADIENT BOOSTING: A PRACTICAL APPROACH

MODELADO DE IMPRESIÓN PREDICTIVO CON SKLEARN Y AUMENTO DE GRADIENTE: UN ENFOQUE PRÁCTICO

535

Mário Sérgio Martinelli Medina^{*}

Rodrigo de Oliveira Gonçalves^{**}

Wellington Vida Leal^{***}

Fernanda Sabião Espinhara^{****}

Rodrigo Henrique Cunha Palácios^{*****}

André Roberto Ortoncelli^{*****}

RESUMO: Este estudo propõe a utilização de técnicas de Inteligência Artificial (IA), implementadas em Python com a biblioteca scikit-learn, para prever a demanda de impressões em uma Instituição Pública de Ensino Superior. Com o objetivo de facilitar processos de licitação para serviços de outsourcing de impressão, foi desenvolvido um modelo de aprendizado de máquina supervisionado baseado em Gradient Boosting, um método de ensemble. Este modelo utiliza dados históricos de impressões para estimar com precisão a quantidade de impressões a serem geradas em períodos futuros. Espera-se que essa abordagem proporcione uma base sólida para a tomada de decisões estratégicas relacionadas à gestão de recursos e contratação de serviços de impressão, otimizando a eficiência operacional e financeira da instituição.

^{*}Especialista em Redes de Computadores. Instituição de formação: Universidade Tecnológica Federal do Paraná. Endereço: (Cornélio Procópio, Paraná, Brasil). E-mail: mariomedina@utfpr.edu.br

^{**}Especialista em Redes de Computadores. Instituição de formação: Universidade Tecnológica Federal do Paraná. Endereço: (Cornélio Procópio, Paraná, Brasil). E-mail: rgoncalves@utfpr.edu.br

^{***}Especialista em Redes de Computadores. Instituição de formação: Universidade Tecnológica Federal do Paraná. Endereço: (Cornélio Procópio, Paraná, Brasil). E-mail: wvleal@utfpr.edu.br

^{****}Especialista em Direito Processual Civil. Instituição de formação: Universidade Estadual do Norte do Paraná. Endereço: (Santa Mariana, Paraná, Brasil). E-mail: fsespinhara@utfpr.edu.br

^{*****}Doutor em Engenharia Elétrica. Instituição de formação: Universidade de São Paulo. Endereço: (Cornélio Procópio, Paraná, Brasil). E-mail: rodrigopalacios@utfpr.edu.br

^{*****}Doutor em Informática. Instituição de formação: Universidade Tecnológica Federal do Paraná. Endereço: (Dois Vizinhos, Paraná, Brasil). E-mail: ortoncelli@utfpr.edu.br

PALAVRAS-CHAVE: Inteligência Artificial, Outsourcing de Impressão, Gradient Boosting, Scikit-Learn, Python.

ABSTRACT: This study proposes the use of Artificial Intelligence (AI) techniques, implemented in Python with the scikit-learn library, to predict printing demand at the Public Higher Education Institution. With the aim of facilitating the bidding process for printing outsourcing services, a supervised machine learning model based on Gradient Boosting, an ensemble method, has been developed. This model utilizes historical printing data to accurately estimate the quantity of prints to be generated in future periods. It is expected that this approach will provide a solid foundation for strategic decision-making related to resource management and printing service procurement, optimizing the operational and financial efficiency of the institution.

KEYWORDS: Artificial Intelligence, Printing Outsourcing, Gradient Boosting, Scikit-Learn, Python.

RESUMEN: Este estudio propone utilizar técnicas de Inteligencia Artificial (IA), implementadas en Python con la biblioteca scikit-learn, para prever la demanda de impresiones en la Institución Pública de Educación Superior. Con el objetivo de facilitar procesos de licitación para servicios de externalización de impresión, se desarrolló un modelo de aprendizaje supervisado basado en Gradient Boosting, un método de conjunto. Este modelo utiliza datos históricos de impresiones para estimar con precisión la cantidad de impresiones que se generarán en períodos futuros. Se espera que este enfoque proporcione una base sólida para la toma de decisiones estratégicas relacionadas con la gestión de recursos y la contratación de servicios de impresión, optimizando la eficiencia operativa y financiera de la institución.

PALABRAS CLAVE: Inteligencia artificial, Subcontratación de Impresión, Gradient Boosting, Scikit-Learn, Python.

1 INTRODUÇÃO

Nos últimos anos, a Inteligência Artificial (IA) tem desempenhado um papel cada vez mais significativo na transformação digital de diversas indústrias e setores, proporcionando avanços notáveis em eficiência operacional, tomada de decisões e previsão de tendências. No contexto da Instituição Pública de Ensino superior em questão, a gestão de recursos públicos sempre prioriza a aplicação do princípio da eficiência e da economicidade e neste sentido, a aplicação da Inteligência Artificial emerge como uma solução promissora para otimizar processos administrativos e operacionais.

Um aspecto de suma importância na gestão universitária é a administração dos serviços de impressão, que desempenham um papel fundamental no suporte às atividades acadêmicas, administrativas e de pesquisa e extensão, pilares estes da instituição. No entanto, a projeção precisa da demanda de impressões é um desafio constante, especialmente ao considerar fatores sazonais, variações nas necessidades de diferentes departamentos e a influência de eventos específicos que podem impactar a demanda de recursos.

Neste cenário, surge a necessidade de desenvolver um modelo de previsão robusto que possa estimar com a maior precisão possível a quantidade de impressões a serem demandadas em períodos futuros. Essa previsão é essencial não apenas para garantir a adequação dos recursos de impressão disponíveis, mas também para auxiliar no processo de licitação para a contratação de serviços de outsourcing de impressão, que seguem uma série de princípios basilares. Uma estimativa precisa da demanda futura permite uma alocação eficiente de recursos financeiros e garante que os serviços contratados atendam às necessidades reais da comunidade universitária, sem que haja discrepância no modelo contratado, evitando desperdício e também retrabalho.

Este artigo propõe a utilização de técnicas de Inteligência Artificial para desenvolver um modelo preditivo que seja capaz de prever as demandas de impressões com base em dados históricos. O objetivo principal é fornecer à Instituição uma ferramenta eficaz para antecipar as necessidades futuras de impressão, permitindo uma gestão eficiente e estratégica dos serviços e recursos de impressão.

Para alcançar esse objetivo, foi empregado um modelo de aprendizado de máquina supervisionado, que é uma abordagem em que um algoritmo é treinado usando um conjunto de dados que inclui exemplos de entradas e as saídas desejadas correspondentes, tendo como objetivo aprender uma função que mapeia as entradas para as saídas, de modo que o modelo possa fazer previsões precisas para novos dados com base no que foi aprendido durante o treinamento, utilizando especificamente o algoritmo de Gradient Boosting, devido à sua capacidade de analisar padrões em conjuntos de dados e fazer previsões com maior exatidão. O modelo foi treinado utilizando uma base de dados pretéritos de impressões geradas

mensalmente, por cada impressora instalada na Instituição em estudo, ao longo de um período de dois anos, incorporando variáveis relevantes que possam influenciar a demanda, como sazonalidade, tendências de uso e eventos específicos.

A implementação bem-sucedida deste modelo preditivo não só permitirá antecipar e responder proativamente às flutuações na demanda de impressões, mas também proporcionará uma base sólida para a tomada de decisões estratégicas relacionadas à gestão de recursos e a contratação de serviços de outsourcing de impressão. Além disso, este estudo contribuirá para ampliar o entendimento sobre o potencial da Inteligência Artificial na otimização de processos administrativos e demonstrará sua relevância na promoção de uma gestão eficiente e sustentável dos recursos financeiros.

2 DEFINIÇÃO DO PROBLEMA

A Instituição Pública de Ensino Superior em tela, encara o desafio de gerenciar seus serviços de impressão de forma eficiente, uma vez que desempenham um papel crucial no suporte às atividades acadêmicas, administrativas e de pesquisa, garantindo a melhor aplicação de recursos financeiros. A precisão na projeção da demanda de impressões é primordial para garantir a adequação dos recursos disponíveis e orientar o processo de licitação, a fim de balizar o quantitativo, para a contratação de serviços de outsourcing de impressão.

Atualmente, conta com vinte e quatro impressoras monocromáticas e duas impressoras policromáticas, todas operadas por meio de um serviço de contrato de outsourcing de impressão firmado por meio de processo licitatório. Durante o planejamento desses contratos, é crucial estimar com o máximo de precisão possível a quantidade de impressões a serem realizadas ao longo do período acordado. Essa estimativa é fundamental para determinar os valores que serão pagos pela franquia contratada. Caso o número de impressões esteja equivocado, poderá haver necessidade de discutir o equilíbrio econômico-financeiro do contrato, que poderão causar inclusive prejuízos à Administração Pública.

O cerne do problema reside na necessidade de prever de forma mais precisa a

demanda futura de gastos com impressões, levando em consideração fatores como sazonalidade, tendências de uso e eventos específicos que possam influenciar nos padrões de utilização. Uma projeção inadequada pode resultar em subutilização ou superutilização dos recursos destinados ao contrato de outsourcing de impressão, afetando negativamente a prestação do serviço.

Portanto, é essencial desenvolver um modelo preditivo robusto capaz de antecipar com precisão as necessidades futuras de impressão na Instituição. Esse modelo não apenas facilitará a tomada de decisões estratégicas relacionadas à gestão de recursos, mas também otimizará os futuros processos de contratação por meio de licitação para serviços de outsourcing de impressão, garantindo que os serviços contratados atendam efetivamente com excelência às demandas reais da comunidade universitária.

3 SOLUÇÃO DO PROBLEMA

Em primeiro momento, os dados foram coletados e organizados em um formato que fosse adequado para serem utilizados nos modelos preditivos estudados. Elaborou-se um arquivo no formato ".CSV" (comma-separated-values), que inclui o registro de todas as impressoras locadas por meio do processo de contratação, instaladas e operacionais, juntamente com a quantidade de impressões realizadas mensalmente, abrangendo o período de fevereiro de 2022 a fevereiro de 2024, conforme exemplificado na Figura 1.

Figura 1. Dados históricos das impressões.

Impressora	01/02/2022	01/03/2022	01/04/2022	...	01/02/2024
COGERH_DCPL5652DN	916	1326	3767	...	702
DEMAP_DCPL5652DN	1217	1045	1739	...	706
DEPRO_DCPL5652DN	103	75	111	...	211
GADIR_DCPL5652DN	532	744	560	...	337
DIREC_DCPL5652DN	192	563	365	...	147
DIRPPG_DCPL5652DN	109	268	722	...	367
SEGEA_DCPL5652DN	401	815	1973	...	268
DERAC_DCPL5652DN	1795	1358	991	...	524
FISICA_DCPL5652DN	3	242	641	...	0
AMBULATORIO_DCPL5652DN	38	33	98	...	36
QUIMICA_DCPL5652DN	0	10	251	...	2
QUADRA_DCPL5652DN	2	15	14	...	139
CIPEMA_DCPL5652DN	1	33	464	...	37
GHI_01_DCPL5652DN	15	637	2261	...	220
GHI_02_DCPL5652DN	2	51	1404	...	272
K_DCPL5652DN	160	58	246	...	173
LINGUAS_DCPL5652DN	0	326	438	...	343
SECCOORD_DCPL5652DN	447	828	1317	...	776
CIPECA_DCPL5652DN	703	601	564	...	236
INCUBADORA_DCPL5652DN	187	431	586	...	78
S323_02_DCPL5652DN	0	26	272	...	577
S323_01_DCPL5652DN	325	483	1421	...	64
G106_DCPL5652DN	0	4	171	...	171
PROFMAT_DCPL5652DN	350	336	356	...	135
DEPRO_MFCT4500DW	1	7	21	...	2
DIREC_MFCT4500DW	209	189	33	...	59

Fonte: Elaborada pelos próprios autores.

Para enfrentar o desafio de uma previsão assertiva de quantidades de impressões, foram explorados diferentes modelos de previsão, incluindo a regressão linear, árvores de decisão e o Gradient Boosting, a fim de identificar a abordagem mais eficaz.

A regressão linear consiste em uma técnica estatística utilizada para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes, sendo essas as variáveis preditoras. Simplificando, ela visa encontrar uma linha reta que melhor se amolda aos dados observados, de forma a prever ou estimar valores futuros da variável dependente com base nas variáveis independentes. Na predição de dados, a regressão linear é frequentemente usada para fazer previsões sobre um

fenômeno com base em dados históricos. Por exemplo, caso hajam dados de vendas de uma empresa ao longo dos últimos meses, juntamente com informações como publicidade, preço do produto, e outras variáveis relacionadas às vendas. Usando a regressão linear, é possível criar um modelo que relacione essas variáveis independentes com as vendas, permitindo prever as vendas futuras com base nas mudanças nas variáveis preditoras. A equação geral de uma regressão linear simples é da forma:

$$y = mx + b$$

onde:

y = variável dependente a ser prevista

m = coeficiente angular da linha de regressão (ou seja, a inclinação da linha)

x = variável independente usada para fazer a previsão

b = coeficiente linear (ou seja, o ponto onde a linha intercepta o eixo vertical)

Em resumo, a regressão linear em predição de dados é uma técnica poderosa para modelar relações entre variáveis e usar esses modelos para fazer previsões sobre dados futuros com base em dados históricos.

As árvores de decisão são modelos de aprendizado de máquina que são amplamente utilizados para previsão e classificação de dados. Elas são chamadas de "árvores" porque assumem uma estrutura semelhante a uma árvore com ramos e folhas. Cada nó interno da árvore representa uma decisão baseada em uma característica dos dados, enquanto as folhas representam as saídas ou resultados finais.

Em termos simples, uma árvore de decisão divide o conjunto de dados em grupos menores com base em diferentes características, buscando separar os dados de uma forma que maximize a pureza ou a homogeneidade de cada grupo em relação à variável de destino (a variável que se busca prever).

Existem diferentes algoritmos para construir árvores de decisão, como o algoritmo CART (Classification and Regression Trees) e o algoritmo ID3 (Iterative

Dichotomiser 3). Esses algoritmos funcionam de forma iterativa, selecionando a melhor característica para dividir os dados em cada nó da árvore, de modo a minimizar a impureza nos grupos resultantes.

As árvores de decisão são úteis em predição de dados porque apresentam:

- a) interpretabilidade: São facilmente compreensíveis e interpretáveis, permitindo que os usuários entendam como as decisões são tomadas;
- b) lidar com dados não lineares: São capazes de lidar com relações não lineares entre as variáveis;
- c) lidar com variáveis categóricas e numéricas: Podem lidar com variáveis categóricas e numéricas sem a necessidade de pré-processamento extensivo;
- d) robustez: São robustas em relação a outliers e dados ruidosos.

No entanto, é importante notar que árvores de decisão podem ser propensas a overfitting (sobreajuste) se não forem devidamente controladas por técnicas como a poda da árvore ou o uso de florestas aleatórias (Random Forests), que são uma extensão das árvores de decisão e trabalham com múltiplas árvores para melhorar a generalização do modelo.

O Gradient Boosting consiste em um poderoso algoritmo de aprendizado de máquina utilizado para realizar tarefas de predição em conjuntos de dados. Ele pertence à categoria de modelos ensemble, que combinam múltiplos modelos mais simples para formar um modelo mais robusto e preciso. Em particular, o Gradient Boosting é uma técnica de boosting, que cria um modelo forte a partir da combinação de vários modelos mais fracos.

A ideia central por trás do Gradient Boosting é treinar modelos sequencialmente, onde cada novo modelo é treinado para corrigir os erros cometidos pelo modelo anterior. Isso é feito de forma iterativa, de modo que o modelo final seja uma combinação ponderada dos modelos individuais, com pesos atribuídos com base na sua contribuição para a redução do erro total.

O processo de Gradient Boosting apresenta:

- a) inicialização: Um modelo inicial (geralmente simples, como uma árvore de decisão) é treinado no conjunto de dados. Este modelo inicial é frequentemente chamado de "modelo base" ou "estimador fraco";
- b) iteração: Para cada iteração, um novo modelo é treinado para capturar os erros residuais do modelo anterior. Isso é feito ajustando os pesos dos exemplos de treinamento de forma que o novo modelo se concentre nas instâncias que foram mal previstas pelo modelo anterior;
- c) atualização dos pesos: Os pesos dos exemplos de treinamento são atualizados com base na diferença entre as previsões do modelo atual e as verdadeiras respostas (rótulos) dos exemplos de treinamento;
- d) combinação dos modelos: O modelo final é uma combinação ponderada de todos os modelos individuais, onde o peso de cada modelo na combinação é determinado pelo seu desempenho em corrigir os erros residuais.

O Gradient Boosting é conhecido por sua capacidade de produzir modelos altamente precisos e robustos, especialmente em problemas complexos e de grande escala. Além disso, ele pode lidar com diferentes tipos de variáveis (categóricas e numéricas) e é menos propenso a overfitting em comparação com outros algoritmos complexos, desde que sejam ajustados adequadamente os hiperparâmetros, como a taxa de aprendizado e a profundidade máxima das árvores. Além disso, variantes populares do Gradient Boosting incluem o XGBoost e o LightGBM, que são implementações otimizadas e eficientes desse algoritmo.

Um ensemble, como citado na exemplificação do Gradient Boosting, é uma técnica em aprendizado de máquina onde múltiplos modelos são combinados para formar um modelo mais forte e geralmente mais preciso do que cada um dos modelos individuais. A ideia por trás de um ensemble é explorar a diversidade dos modelos e combinar suas previsões para obter um desempenho melhor do que qualquer modelo individualmente.

Existem duas abordagens principais para criar ensembles: o bagging e o boosting, sendo que este último foi o utilizado no presente trabalho.

Após uma análise comparativa entre os modelos, optou-se pelo uso do algoritmo de Gradient Boosting devido à sua capacidade de lidar com relações não lineares entre variáveis e de capturar padrões complexos nos dados históricos de impressões.

Os algoritmos baseados em Árvores são modelos vastamente utilizados em aprendizado supervisionado, pois possuem desempenho satisfatório em problemas não lineares. São adaptáveis para qualquer tipo de problema, tanto regressão quanto classificação. São largamente utilizados em problemas ligados à Ciência de Dados. Porém, está distante de ser uma ferramenta ideal para o aprendizado preditivo, devido a sua imprecisão. Para contornar, é utilizado método denominado ensemble, como exemplificado acima, que envolve um grupo de modelos de predição para obter uma melhor acurácia e estabilidade do modelo, através de decisões de vários classificadores para melhorar o desempenho do modelo como um todo. Aqui entram os modelos Boosting.

O Gradient Boosting consiste em um método de aprendizado de máquina para problemas de regressão e classificação, que produz um modelo de previsão na forma de um ensemble de modelos de previsão fracos, com o objetivo de produzir um melhor modelo preditivo. Ele constrói o modelo em etapas, como outros métodos de reforço e os generaliza, permitindo a otimização de uma função de perda diferencial arbitrária. Os preditores fracos são modelos que se utilizados isoladamente, possuem uma acurácia abaixo do valor esperado.

Nesta técnica, cada classificador fraco é treinado com um conjunto de dados sequenciais e adaptativos, onde um modelo base depende dos anteriores e ao final, são combinados de maneira determinística. O objetivo de cada modelo fraco, é minimizar o erro do modelo anterior por meio de uma função de perda. Árvore Boosting é a combinação de modelos simples, estes são chamados de modelos fracos. Tipicamente esses modelos são Árvores de Decisão, o algoritmo combina classificadores fracos com o intuito de produzir um classificador forte (Hastie; Tibshirani; Friedman, 2001).

Essa abordagem faz uso de modelos aditivos para gradualmente aproximar um melhor modelo, de modo a somar submodelos ao modelo composto. Árvores de

Decisão tendem a gerar sobreajuste (overfitting), e para solucionar este problema, o Gradient Boosting é implementado (Zheng; Casari, 2018).

Ao utilizar o algoritmo de Gradient Boosting, uma biblioteca presente no conjunto de ferramentas do Python, conhecido como sklearn, que oferece uma ampla gama de ferramentas para tarefas de aprendizado supervisionado e não supervisionado, bem como para pré-processamento de dados e avaliação de modelos, foi possível aprimorar significativamente as previsões de demanda de impressões. O objetivo principal do scikit-learn é fornecer uma interface simples e eficiente para desenvolver e implementar modelos de aprendizado de máquina. O Gradient Boosting é então, uma técnica de aprendizado de máquina que constrói um modelo preditivo por meio da combinação de várias previsões mais simples, chamadas de "weak learners", em um único modelo robusto. Durante o treinamento, o algoritmo faz iterações sucessivas sobre os dados, ajustando cada modelo subsequente para corrigir os erros cometidos pelos modelos anteriores. Essa abordagem permite que o modelo aprenda com os erros e se adapte continuamente (Géron, 2019), resultando em previsões mais precisas e confiáveis.

O modelo escolhido utiliza dados pretéritos, considerando métricas como volume de impressões, sazonalidade e tendências de uso, para antecipar com precisão as necessidades futuras de impressão em períodos específicos. Além disso, o Gradient Boosting faz uso dos dados previstos em cada iteração do treinamento, permitindo uma adaptação contínua do modelo às mudanças na demanda ao longo do tempo. Tais características permitem tomar decisões mais certas em relação à gestão de recursos de impressão, garantindo uma alocação eficiente de recursos e otimizando o uso dos recursos financeiros.

4 RESULTADOS DA IMPLEMENTAÇÃO

Os resultados gerados, conforme mostrado de forma suprimida na Figura 2, foram obtidos através da implementação de códigos de programação em linguagem Python e suas bibliotecas, como Pandas, Numpy, Sklearn, GradientBoostingRegressor entre outras, permitindo uma avaliação preliminar dos

resultados obtidos após a implementação. Com as informações contidas na base histórica, de fevereiro de 2022 à fevereiro de 2024, o modelo implementado gerou a previsão, mensalmente, até dezembro de 2030.

Figura 2. Previsão gerada pelo algoritmo.

Impressora	01/02/2022	...	01/12/2024	...	01/12/2030
COGERH_DCPL5652DN	916	...	1028	...	976
DEMAP_DCPL5652DN	1217	...	988	...	1093
DEPRO_DCPL5652DN	103	...	157	...	157
GADIR_DCPL5652DN	532	...	661	...	741
DIREC_DCPL5652DN	192	...	276	...	273
DIRPPG_DCPL5652DN	109	...	163	...	237
SEGEA_DCPL5652DN	401	...	520	...	565
DERAC_DCPL5652DN	1795	...	614	...	580
FISICA_DCPL5652DN	3	...	70	...	83
AMBULATORIO_DCPL5652DN	38	...	172	...	155
QUIMICA_DCPL5652DN	0	...	104	...	143
QUADRA_DCPL5652DN	2	...	187	...	207
CIPEMA_DCPL5652DN	1	...	273	...	337
GHI_01_DCPL5652DN	15	...	410	...	976
GHI_02_DCPL5652DN	2	...	191	...	555
K_DCPL5652DN	160	...	92	...	294
LINGUAS_DCPL5652DN	0	...	210	...	219
SECCOORD_DCPL5652DN	447	...	935	...	1046
CIPECA_DCPL5652DN	703	...	715	...	729
INCUBADORA_DCPL5652DN	187	...	235	...	262
S323_02_DCPL5652DN	0	...	141	...	121
S323_01_DCPL5652DN	325	...	785	...	849
G106_DCPL5652DN	0	...	51	...	112
PROFMAT_DCPL5652DN	350	...	34	...	49
DEPRO_MFCT4500DW	1	...	18	...	21
DIREC_MFCT4500DW	209	...	123	...	129
Total	7708	...	9153	...	10909
Total por ano			192213		214858

Fonte: Elaborada pelos próprios autores.

Durante o processo de implementação e teste do modelo de previsão de demanda de impressões baseado em Gradient Boosting foram observados resultados promissores, embora ainda preliminares, que merecem maior dedicação de tempo para testes do modelo. As previsões geradas pelo modelo mostraram-se compatíveis com as expectativas iniciais, fornecendo estimativas que refletiam de forma satisfatória os padrões históricos de demanda por impressões.

Ao analisar os resultados preliminares obtidos, foi possível identificar que o modelo demonstrou uma capacidade significativa de capturar padrões complexos nos dados históricos de impressões. Isso sugere que o Gradient Boosting é uma ferramenta eficaz para lidar com relações não lineares entre variáveis e para antecipar de forma precisa as necessidades futuras de impressão em períodos específicos.

No entanto, é importante ressaltar que o uso do modelo ainda está em estágio inicial e está sujeito a aprimoramentos contínuos a fim de identificar possíveis discrepâncias com amostragens maiores de dados. Deve ser realizada uma análise mais aprofundada com a inclusão de novos dados, buscando a validação e refino do desempenho do modelo. Essa análise detalhada permitirá uma compreensão mais completa do potencial impacto da implementação do modelo na gestão de recursos empregados nos contratos de outsourcing de impressão.

Além disso, deverá ser realizada uma avaliação mais abrangente dos resultados, levando em consideração não apenas a precisão das previsões, mas também outros aspectos importantes, como a eficiência operacional, a otimização de recursos e os custos associados. Essa análise holística permitirá avaliar de forma mais abrangente o valor agregado pelo modelo e sua viabilidade para aprimorar os processos de gestão de recursos na administração pública.

Em suma, os resultados iniciais são encorajadores e indicam que o modelo de previsão baseado em Gradient Boosting tem o potencial de fornecer insights valiosos para aprimorar a gestão de recursos de impressão. No entanto, é necessário continuar monitorando e refinando o modelo para garantir sua eficácia e relevância contínua no contexto operacional da Instituição.

5 ANÁLISE DOS RESULTADOS

Após a geração das previsões de demanda de impressões utilizando o modelo baseado em Gradient Boosting, foi realizada uma análise detalhada, embora preliminar, dos resultados obtidos. Nesta seção, serão destacadas algumas das principais observações derivadas dessa análise.

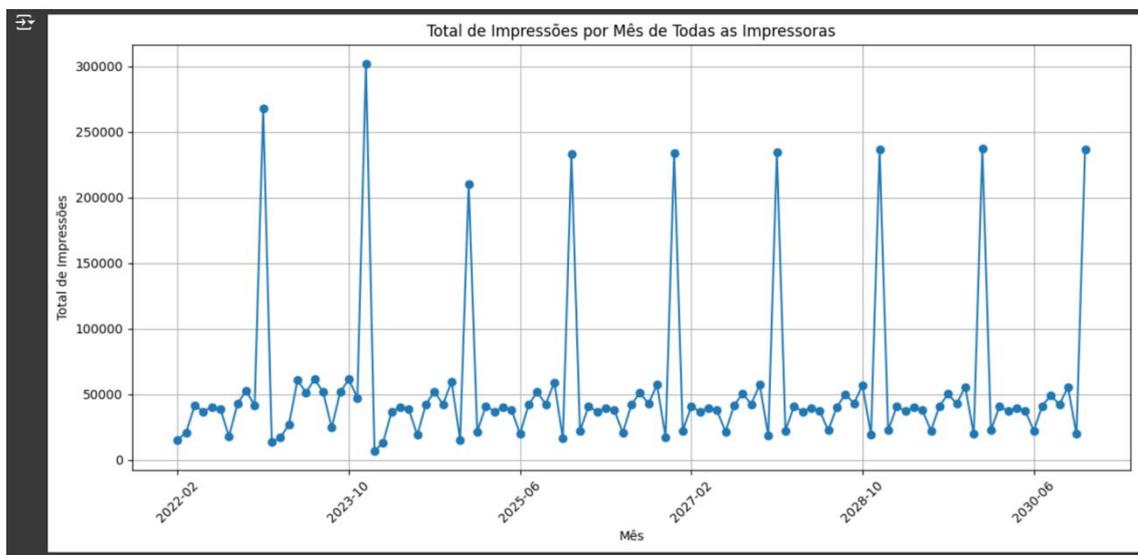
Primeiramente é possível notar que as previsões geradas pelo modelo apresentaram uma correspondência satisfatória com os padrões históricos de demanda por impressões. Isso indica que o modelo foi capaz de capturar efetivamente as tendências e sazonalidades presentes nos dados históricos, fornecendo estimativas que refletem de forma precisa as variações esperadas na demanda ao longo do tempo, trazendo grande coerência no resultado.

Além disso, ao comparar as previsões geradas pelo modelo com os dados reais de impressões, observa-se que o modelo demonstra uma capacidade consistente de antecipar as flutuações na demanda de impressões. Em muitos casos, as previsões estavam alinhadas de perto com os valores reais observados, indicando uma boa precisão do modelo em prever as necessidades futuras de impressão.

Também é possível identificar algumas áreas que podem ser alvo de aprimoramento. Em certos períodos ou situações específicas, o modelo apresentou desvios significativos entre as previsões e os dados reais de impressões. Esses desvios podem ser atribuídos a uma série de fatores, como eventos inesperados, mudanças nas condições operacionais ou limitações nos dados disponíveis para treinamento do modelo.

Na Figura 3, ficam mais evidentes tais constatações:

Figura 3. Gráfico da previsão do total de impressões.



Fonte: Elaborada pelos próprios autores.

O gráfico ilustra a quantidade total de impressões realizadas por todas as impressoras ao longo dos meses. Cada ponto no gráfico representa o total de impressões somadas de todas as impressoras em um determinado mês.

A análise do gráfico revela tendências e variações mensais no uso das impressoras, o que pode ser útil para identificar padrões de uso e prever futuras demandas. Por exemplo, picos em certos meses podem indicar períodos de maior atividade, possivelmente ligados a ciclos acadêmicos ou administrativos.

A linha de tendência permite visualizar a variação mensal e identificar períodos de maior ou menor utilização. Esse tipo de análise é fundamental para o planejamento e a gestão eficiente dos recursos de impressão, ajudando a otimizar o uso das impressoras e a planejar manutenções preventivas e substituições de suprimentos.

A identificação de padrões sazonais pode ajudar a prever necessidades futuras e ajustar os serviços de impressão conforme a demanda, garantindo um melhor atendimento às necessidades institucionais.

Diante dessas observações, fica evidente a importância de uma abordagem contínua e iterativa para o aprimoramento do modelo de previsão de demanda de impressões. Pretende-se realizar análises mais aprofundadas para identificar as causas subjacentes dos desvios observados e implementar ajustes no modelo, conforme necessário, para melhorar sua precisão e robustez.

Além disso, é indicado explorar métodos adicionais de validação do modelo, como a divisão dos dados em conjuntos de treinamento e teste, a fim de garantir uma avaliação mais rigorosa de seu desempenho e generalização para novos dados.

Em resumo, a análise dos resultados destaca tanto os pontos fortes quanto as áreas de melhoria do modelo de previsão de demanda de impressões baseado em Gradient Boosting. Esses insights fornecerão orientações valiosas para refinamentos e aprimoramentos futuros do modelo, visando melhorar sua capacidade de antecipar com precisão as necessidades futuras e apoiar a gestão eficaz dos recursos de impressão.

6 CONCLUSÕES

Os testes iniciais do modelo de previsão de demanda de impressões baseado em Gradient Boosting representam um passo significativo em direção à melhoria da gestão de recursos de impressão na Instituição de Ensino Superior estudada. Embora o modelo ainda esteja em fase inicial de testes e validações, os resultados indicam um potencial promissor para aprimorar a capacidade de antecipar e atender às demandas futuras por serviços de impressão, de modo a se obter a racionalização dos recursos e maior assertividade no momento de realizar o processo licitatório.

Mesmo que ainda não tenha sido implementado operacionalmente, o modelo fornece uma base sólida para análises futuras e pode servir como um valioso instrumento de apoio à tomada de decisões relacionadas à contratação de serviços de outsourcing de impressão. Ao capturar padrões complexos nos dados históricos de impressões e adaptar dinamicamente suas previsões às mudanças na demanda, o modelo oferece dados valiosos para otimizar a alocação de recursos e garantir uma operação mais eficiente e sustentável da Instituição, eficiência esta, que é um dos pilares da administração pública.

À medida que novos avanços são alcançados nesse processo, pode-se observar que o modelo de previsão de demanda de impressões baseado em Gradient Boosting se tornará uma ferramenta essencial para auxiliar os gestores da Instituição de Ensino na tomada de decisões estratégicas relacionadas à gestão de recursos de impressão e na otimização dos processos de licitação para serviços de outsourcing.

REFERÊNCIAS

- BEAZLEY, D. M. **Python Essential Reference**. Addison-Wesley Professional, 2009.
- GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**. O'Reilly Media, 2019.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Interference and prediction**. Springer: 2nd ed., 2001.



<https://www.faccrei.edu.br/revista>

KE, G.; MENG, Q.; FINLEY, T.; WANG, T.; CHEN, W.; MA, W.; YE, Q.; LIU, T. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." **Advances in Neural Information Processing Systems**, vol. 30, 2017.

RASCHKA, S.; MIRJALILI, V. **Python Machine Learning**: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Packt, 2019.

VANDERPLAS, J. **Python Data Science Handbook**: Essential Tools for Working with Data. O'Reilly Media, 2016.

ZHENG, A.; CASARI, A. **Feature Engineering for Machine Learning**: Principles and Techniques for Data Scientists. O'Reilly Media, 2018.

Recebido em: 11/07/2024.

Aprovado em: 08/08/2024.