



Universidade Federal de Ouro Preto
Escola de Minas
CECAU - Colegiado do Curso de
Engenharia de Controle e Automação



Lucas de Paula Moreira

**Dashboard para gerenciamento e controle financeiro com previsão
através de inteligência artificial**

Monografia de Graduação

Ouro Preto, 2024

Lucas de Paula Moreira

Dashboard para gerenciamento e controle financeiro com previsão através de inteligência artificial

Trabalho apresentado ao Colegiado do Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como parte dos requisitos para a obtenção do Grau de Engenheira(o) de Controle e Automação.

Universidade Federal de Ouro Preto

Orientador: Profa. Dra. Adrielle de Carvalho Santana

Ouro Preto

2024



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
ESCOLA DE MINAS
DEPARTAMENTO DE ENGENHARIA CONTROLE E
AUTOMACAO



FOLHA DE APROVAÇÃO

Lucas de Paula Moreira

Dashboard para gerenciamento e controle financeiro com previsão através de inteligência artificial

Monografia apresentada ao Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de bacharel em Engenharia de Controle e Automação

Aprovada em 22 de fevereiro de 2024

Membros da banca

Dra. Adrielle de Carvalho Santana - Orientadora (Universidade Federal de Ouro Preto)
Dr. Agnaldo José da Rocha Reis - Convidado (Universidade Federal de Ouro Preto)
Dr. Marcus Vinícius Marques de Paula - Convidado (Accenture)

Adrielle de Carvalho Santana, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 23/02/2024



Documento assinado eletronicamente por **Adrielle de Carvalho Santana, PROFESSOR DE MAGISTERIO SUPERIOR**, em 23/02/2024, às 08:48, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0670684** e o código CRC **A061D2F6**.

Agradecimentos

Gostaria de expressar minha profunda gratidão a todos que tornaram possível a conclusão deste trabalho de conclusão de curso:

À minha família, cujo apoio incondicional, encorajamento, compreensão e sacrifícios foram fundamentais neste passo em busca da graduação.

Aos meus amigos, tanto aqueles que estiveram ao meu lado desde o início quanto os que conheci ao longo desse percurso desafiador. Compartilharam comigo momentos de dificuldade e alegria, e suas palavras de incentivo e apoio moral foram essenciais para superar os obstáculos desta jornada.

À Adrielle, minha orientadora, agradeço por sua ajuda, paciência e dedicação, que desempenharam um papel fundamental no desenvolvimento deste trabalho.

À Escola de Minas de Ouro Preto e à UFOP, expresso meu reconhecimento pela infraestrutura e recursos disponíveis, que desempenharam um papel crucial na realização deste trabalho. Agradeço também aos professores e funcionários pelo ambiente propício ao aprendizado.

Quero expressar minha gratidão a todos que, de alguma forma, contribuíram para este trabalho, mesmo que não tenham sido mencionados especificamente. Cada colaboração foi valiosa e contribuiu para o sucesso deste projeto.

Este trabalho representa não apenas meu esforço individual durante minha graduação, mas sim uma colaboração coletiva de muitas pessoas. Mais uma vez, obrigado a todos que tornaram possível a conclusão desta etapa.

Resumo

O Controle Financeiro tem o objetivo de auxiliar as pessoas a gerirem melhor sua qualidade de vida, permitindo um maior conhecimento da sua situação financeira e compreensão das necessidades para o alcance dos objetivos ao longo da vida. Um método para explorar melhor a tecnologia e auxiliar nesse tópico é utilizar um *dashboard* ou portal para o gerenciamento das finanças, onde cada usuário tenha um detalhamento mais preciso da sua situação pessoal. Neste trabalho, é desenvolvida uma aplicação com o auxílio do *Microsoft Office* e suas ferramentas para o gerenciamento e controle financeiro. Esta aplicação fornece não apenas resumos e montantes de gastos por período, mas também previsões de faturas futuras do usuário. Além disso, utiliza a base de dados do usuário para aplicar modelos de aprendizado de máquina, prevendo gastos futuros de acordo com as categorias de dados presentes no dia a dia do usuário. Este trabalho treinou modelos de *Random Forest* (Floresta Aleatória) e Redes Neurais para a previsão categorizada de dados. O modelo que apresentou o melhor desempenho preditivo foi escolhido para ser utilizado no *dashboard*. Os resultados mostraram que o *Random Forest* obteve um melhor desempenho preditivo para a base de dados utilizada. A partir desses resultados, discutiram-se possíveis abordagens para integrar a previsão de dados ao *dashboard* desenvolvido. Neste trabalho, é mostrado não apenas como a utilização de uma aplicação para gestão financeira é importante e pode ajudar o usuário nessa tarefa, mas também aborda a possibilidade de utilização de modelos baseados em inteligência artificial no setor financeiro.

Palavras-chaves: aprendizado de máquina, controle financeiro, *dashboard*, *random forest*, redes neurais.

Abstract

Financial control aims to assist people in managing their quality of life better, providing a deeper understanding of their financial situation and insight into the needs to achieve their goals throughout life. A method to leverage technology for better assistance in this matter is to use a dashboard or portal for financial management, where each user has a more detailed overview of their personal situation. In this work, an application is developed with the assistance of Microsoft Office and its tools for financial management and control. This application not only provides summaries and amounts of expenses per period but also forecasts future user invoices. Additionally, it uses the user's database to apply machine learning models, predicting future expenses based on the data categories present in the user's daily life. This work trained Random Forest and Neural Network models for categorized data prediction. The model that showed the best predictive performance was chosen for use in the dashboard. The results indicated that Random Forest outperformed in predictive performance for the utilized database. From these results, possible approaches to integrating data prediction into the developed dashboard were discussed. This work demonstrates not only how the use of an application for financial management is important and can assist the user in this task but also addresses the possibility of using artificial intelligence-based models in the financial sector.

Key-words: financial control, machine learning, dashboard, random forest, neural network.

Lista de ilustrações

Figura 1 – Trecho de fórmula da base de dados	24
Figura 2 – Fluxograma de processos do projeto	27
Figura 3 – Trecho onde os dados do usuário são inseridos para serem salvos na base de dados.	28
Figura 4 – Tabelas base para preenchimento das transações financeiras no Excel. .	28
Figura 5 – Trecho da planilha com busca de lançamentos do cartão	29
Figura 6 – Seletor de dados da planilha de lançamento de cartões.	29
Figura 7 – Planilha que reúne a visão simplificada dos dados do usuário.	30

Lista de tabelas

Tabela 1 – Hiperparâmetros do Modelo <i>Random Forest</i>	25
Tabela 2 – Hiperparâmetros da Rede Neural.	26
Tabela 3 – Resultados do Modelo <i>Random Forest</i>	30
Tabela 4 – Resultados do Modelo de Rede neural	31
Tabela 5 – Comparação de Resultados entre Random Forest e Rede Neural	31

Lista de abreviaturas e siglas

SPC	Serviço de Proteção ao Crédito
ER	Entidade e Relacionamento
MER	Modelos de Entidade e Relacionamento
TI	Tecnologia da Informação
SI	Sistemas de Informação
VBA	Virtual Basic for Applications

Sumário

1	INTRODUÇÃO	12
1.1	Contextualização	12
1.2	Objetivos	13
1.2.1	Objetivos gerais	13
1.2.2	Objetivos específicos	13
1.3	Metodologia	13
1.4	Organização do texto	15
2	REVISÃO BIBLIOGRÁFICA	16
2.1	Sistemas da informação	16
2.2	Trabalhos relacionados	16
2.3	Aprendizado de máquina	16
2.4	Floresta aleatória	17
2.4.1	Hiperparâmetros de ajuste da Floresta aleatória	17
2.4.1.1	Número de Árvores (<i>n estimators</i>)	18
2.4.1.2	Profundidade Máxima das Árvores (<i>max depth</i>)	18
2.4.1.3	Número Mínimo de Amostras em Folhas (<i>min samples leaf</i>)	18
2.4.1.4	Número Mínimo de Amostras para Divisão (<i>min samples split</i>)	18
2.4.1.5	Número de Recursos a Considerar em Cada Divisão (<i>max features</i>)	18
2.5	Redes neurais	19
2.5.1	Hiperparâmetros de ajuste de Redes Neurais	19
2.5.1.1	Número de Camadas e Neurônios por Camada	19
2.5.1.2	Função de Ativação	19
2.5.1.3	Algoritmo de Otimização	20
2.5.1.4	<i>Batch Size</i>	20
2.5.1.5	Épocas	20
2.6	Métricas de desempenho de machine learning	20
2.6.1	Erro Absoluto Médio (MAE)	20
2.6.2	Erro Quadrático Médio (MSE)	21
2.6.3	Coefficiente de Determinação (R^2)	21
3	DESENVOLVIMENTO	23
3.1	Planejamento da aplicação	23
3.2	Dashboard excel	23
3.3	Manipulação de dados para os modelos de previsão	24
3.4	Treinamento do modelo Random Forest	25

3.5	Treinamento do modelo baseado em Rede Neural	26
3.6	Fluxo de dados	26
4	RESULTADOS	28
4.1	Planilhas de visualização	28
4.2	Análise dos resultados do modelo de Random Forest	30
4.3	Análise dos resultados do modelo de Redes Neurais	30
4.4	Comparativo entre os modelos	31
4.5	Problemas ao utilizar os dados previstos	31
5	CONCLUSÕES	33
5.1	Trabalhos futuros	33
	Referências	35

1 Introdução

1.1 Contextualização

A utilização de sistemas de informação para colaborar com situações do dia-a-dia está cada vez mais visível. De acordo com [Turban, mclean e Wetherbe \(2004\)](#), um sistema de informação coleta, processa, armazena, analisa e dissemina informações com um determinado objetivo. Analisando o momento atual pode-se perceber o mercado voltado a utilizar cada vez mais ferramentas para análise de dados, visando melhor gestão e resultados, uma vez que com os avanços da tecnologia pode-se processar e analisar cada vez mais dados, e como citado por [Robbins \(2000\)](#), a qualidade, bem como a rapidez da tomada de decisão em uma empresa, estão ligadas, fundamentalmente, à qualidade e à disponibilidade das informações.

Ligado ao crescimento de utilização de sistemas de informação para análise de dados, temos o uso do aprendizado de máquina (ML, no inglês *Machine Learning*), que é um conjunto de técnicas de Inteligência Artificial que podem ser utilizadas para reconhecimento de padrões a partir de um conjunto de dados e fazer previsões, análises e classificações não só destes, mas de dados futuros. Como exemplo pode-se citar o trabalho de [Yoshinaga e Castro \(2023\)](#), que aponta como a utilização de Inteligência Artificial impacta o mundos das finanças, sendo tanto em investimento quanto na gestão financeira.

Seguindo no ramo de finanças, seria interessante desenvolver um *dashboard* utilizando Excel/VBA para gerenciamento e controle financeiro, que tem como função auxiliar no controle e análise da saúde financeira do usuário. No trabalho de [Zulian \(2013\)](#), dentro da grande gama de programas disponíveis na área de informática para ajudar no gerenciamento de dados, está o Excel e o VBA. Seu trabalho mostrou também o desenvolvimento e a aplicabilidade destas ferramentas para tarefas rotineiras de uma micro empresa.

Segundo [Macedo \(2007\)](#), O controle financeiro tem como objetivo alcançar a satisfação pessoal, permitindo controle da sua situação financeira e atendimento das necessidades para o alcance dos objetivos ao decorrer da vida, ou ainda podemos dizer que é o controle de suas fontes de renda e despesas, uma vez que para se começar a ter uma saúde financeira você precisa primeiro controlar seus gastos. Um levantamento realizado em 2021 pela Confederação Nacional de Dirigentes Lojistas (CNDL) e pelo Serviço de Proteção ao Crédito (SPC), disponível em [XPEducação \(2021\)](#), mostra que 48% dos consumidores brasileiros não controlam o próprio orçamento.

Muito tem sido discutido sobre a inclusão da educação financeira no ensino fundamental e médio, sendo reforçado todo ano por pesquisas referente ao endividamento da

população brasileira. Como exemplo a pesquisa Divulgada no [G1 \(2022\)](#) onde 77,7% das famílias fecharam o mês de abril de 2022 com dívidas, ultrapassando os meses anteriores. Mas como tal assunto não é obrigatório nas escolas, muito da educação financeira deve partir da própria vontade do cidadão, onde encontra-se então a oportunidade de desenvolver um sistema para auxiliar neste controle de forma inteligente, utilizando as tecnologias atuais.

1.2 Objetivos

1.2.1 Objetivos gerais

Desenvolver um *dashboard* para auxiliar na gestão financeira do usuário, incluindo análise de previsão de gastos por meio de técnicas de aprendizado de máquina utilizando os modelos de floresta aleatória e redes neurais artificiais.

1.2.2 Objetivos específicos

- Mapear os dados que serão requisitados pelo sistema, para criação do banco de dados e suas tabelas;
- Construir e implementar a aplicação;
- Preparar o conjunto de dados para o treinamento dos modelos de aprendizado de máquina;
- Treinar, ajustar e comparar a eficácia da previsão dos modelos;
- Verificar a eficácia da aplicação em relação a análise e previsões futuras.

1.3 Metodologia

O propósito deste trabalho foi utilizar algumas das ferramentas do Excel/VBA para criar um ambiente em que o usuário consiga cadastrar seus dados financeiros, sejam eles receitas, despesas, investimentos entre outras categorias e a partir disso, gerar relatórios, *dashboards* e previsões utilizando um modelo treinado com os dados do usuário para auxiliar em sua saúde financeira.

O primeiro passo da abordagem foi mapear as informações que possam ser relevantes para a aplicação. A princípio tem-se basicamente as categorias classificadas como receitas e despesas, mas para criar uma análise aprofundada podem-se incluir divisões nestas classificações, podendo ser elas despesas fixas, imprevistos, investimentos, reserva de emergência, onde podemos incluir na análise percentuais lógicos baseado em estudos

sobre quanto esta pessoa é aconselhada a gastar, qual faixa é recomendado reservar para emergências, analisar se a pessoa está endividada ou auxiliar a mesma a sair da dívida a partir das análises futuras de seus dados, entre outras opções.

A partir do levantamento de informações, como primeiro passo para criação de aplicações que envolvem uma base de dados, foram criadas as tabelas da base de dados. Para o Excel, pode-se criar a base de dados definindo as colunas e formatando os dados, e depois transformando-os em tabelas, para que estas informações sejam utilizadas para gerar análises a partir da ferramenta tabela dinâmica e outras que serão utilizadas ao longo do projeto.

Criado o modelo da base de dados, inicia-se o desenvolvimento dos campos de inserção e visualização dos dados por parte do usuário, que devem ter uma aparência e identificação intuitiva para o usuário.

O próximo passo foi utilizar as ferramentas de VBA e tabela dinâmica do Excel para criar os filtros e gerenciar a visualização dos dados no *dashboard* de análise. Esta ferramenta permite que a análise de diversos meses, categorias e outros filtros seja visualizada de forma instantânea, somente modificando o filtro na tabela dinâmica.

Seguindo, foi necessário realizar a manipulação da base de dados do usuário para realizar o treinamento dos modelos de aprendizado de máquina, onde foi necessário juntar os parametros dos mesmos até encontrarmos resultados viáveis para a previsão de dados ser efetiva. Os códigos contendo a manipulação e estruturação dos dados ¹, teste do modelo de *Random forest* ² e modelo de Rede Neural ³ estão disponíveis no *Github*. Por fim, foi unificado na aplicação as análises adquiridas pelos modelos de aprendizado de máquina, para que esses possam ser analisados para se verificar se há uma relevância nas informações adquiridas pelo software, para então analisarmos a viabilidade da sua utilização com o propósito de auxiliar na saúde financeira do usuário.

¹ <https://github.com/LucasMoreira7/FinanceAppDatabases/blob/main/TratamentoBaseDeDados.ipynb>

² <https://github.com/LucasMoreira7/FinanceAppDatabases/blob/main/RandomForestFinance.ipynb>

³ <https://github.com/LucasMoreira7/FinanceAppDatabases/blob/main/RedeNeuralFinance.ipynb>

1.4 Organização do texto

No capítulo 1 foi apresentada a contextualização, motivação e objetivos do trabalho. No capítulo 2 é apresentada a revisão bibliográfica que apresenta os conceitos e ferramentas que foram utilizadas para a confecção do mesmo. No capítulo 3 tem-se o desenvolvimento do trabalho, em que se mostra o passo a passo utilizado para desenvolver o *Dashboard* para inserção de dados, o tratamento dos dados para criar o base de treinamento dos modelos, e o treino de cada um dos modelos utilizados. No capítulo 4 são analisados os resultados obtidos durante o trabalho, e no capítulo 5 é descrita a conclusão final junto com idéias para trabalhos futuros.

2 Revisão bibliográfica

2.1 Sistemas da informação

Segundo [Baltzan e Phillips \(2012\)](#), a tecnologia da informação (TI) é um campo dedicado no uso da tecnologia no gerenciamento e no processamento da informação, o que pode ser um importante facilitador do sucesso e da inovação dos negócios. Assim, o termo Sistema de Informações evoluiu de sistema de informações gerenciais (SIG), que é um nome comum para funções de negócios e de disciplinas acadêmicas que abrangem a aplicação de pessoas, tecnologias e procedimentos. Visando realizar a função do SIG, hoje praticamente toda organização possui um setor específico para tratar tal assunto, comumente chamado de Tecnologia da informação (TI) ou Sistemas de Informação (SI) ([BALTZAN; PHILLIPS, 2012](#)). Pode-se diferenciar dados, que são fatos brutos que descrevem características de um evento, de informações, que são dados convertidos em contexto significativo e útil ([BALTZAN; PHILLIPS, 2012](#)).

Presente nos sistemas organizacionais, o uso correto da tecnologia pode estar diretamente ligada ao sucesso e resultados. Como mencionado em [Robbins \(2000\)](#), a qualidade, bem como a rapidez da tomada de decisão em uma empresa, estão ligadas, fundamentalmente, à qualidade e à disponibilidade das informações, o que vem sendo cada vez mais aperfeiçoado em aplicações utilizadas por empresas e indústrias, onde cada vez mais as soluções oferecem resultados mais rápidos, analisando cada vez mais dados.

2.2 Trabalhos relacionados

Como exemplos de aplicações e *softwares* que possuem o mesmo intuito de auxiliar na análise e gestão financeira, temos o *mobills* ([TORO, s.d.](#)), que em seu aplicativo pode-se integrar suas contas bancárias e cartões de crédito, fazendo com que você tenha total controle dos seus gastos de forma automática e a plataforma *Status invest* ([STATUSINVEST... , s.d.](#)), que realiza a gestão e análise de ativos financeiros utilizando a inteligência artificial. O diferencial proposto neste trabalho foi de integrar a gestão de dados financeiros e a previsão através de aprendizado de máquina, proposta não encontrada em outras aplicações.

2.3 Aprendizado de máquina

Para [Bishop \(2006\)](#), aprendizado de máquina é um campo da inteligência artificial focado no desenvolvimento de algoritmos e técnicas que permitem aos computadores aprenderem a partir de dados. O objetivo é capacitar as máquinas para melhorar seu desempenho em

tarefas específicas ao longo do tempo, sem serem explicitamente programadas. Em vez de seguir instruções precisas, os sistemas de aprendizado de máquina são projetados para identificar padrões e regularidades nos dados de treinamento, permitindo que eles façam previsões ou tomem decisões em novos dados. Os modelos de aprendizado de máquina podem ser divididos em dois principais grupos: os supervisionados, em que a partir de dados pré categorizados disponíveis, o modelo é treinado a fim de prever uma variável alvo em que se tem valores para comparação em seu conjunto de dados; e os não supervisionados, em que o modelo não tem dados categorizados para seu aprendizado, ficando por conta do modelo encontrar a relação entre os dados disponíveis.

Como explicado por [Quintino et al. \(2020\)](#), algoritmos supervisionados são utilizados para duas principais de tarefas: tarefas de regressão, quando se deseja prever valores contínuos, enquanto nas tarefas de classificação os algoritmos são usados para atribuir objetos a categorias específicas. Ambas as categorias de algoritmos desempenham um papel crucial no aprendizado de máquina e têm diversas aplicações em diferentes áreas, desde finanças até medicina e reconhecimento de padrões.

2.4 Floresta aleatória

A floresta aleatória (*Random forest*) é um algoritmo construído a partir de um conjunto de árvores de decisão (*Decision Trees*). Na árvore de decisão o algoritmo cria uma série de condições atreladas a tomada de decisões, como em um fluxograma, onde em cada nó do mesmo é verificada uma condição. Se a condição for atendida, o ramo é seguido, caso contrário é seguido outro ramo para análise. Diante disso, pode-se dizer que o *Random forest* é uma melhoria do algoritmo das árvores de decisão, pois, ao invés de utilizar apenas uma árvore de decisão, utilizam-se várias árvores ([GHOSH; JANA; SANYAL, 2019](#)). O modelo é bastante utilizado em diversos ramos, entre eles, a análise de comportamento financeiro que será um alvo deste trabalho. Um exemplo analisado é o trabalho de [Avelar et al. \(2022\)](#) onde utiliza o algoritmo para previsão de comportamento de preço de ativos a partir de dados históricos, e mostra que o modelo foi capaz de obter retornos superiores às médias do mercado em geral.

2.4.1 Hiperparâmetros de ajuste da Floresta aleatória

Não só no *Random Forest*, mas em demais modelos utilizados para realizar tarefas de classificação ou regressão, o ajuste correto dos hiperparâmetros que determinarão não só o método de treinamento, mas também as previsões do modelo são de extrema importância, sendo muitas vezes a parte mais complexa de um treinamento, onde é recomendado o teste do modelo com diversas combinações de hiperparâmetros para que se encontre uma melhor configuração para seu conjunto de dados. Ajustes nos hiperparâmetros não só

ajudam a se treinar um modelo que melhor represente os dados, mas também ajudam a evitar o problema de *overfitting* (sobreajuste) do modelo, que é um fenômeno comum em aprendizado de máquina, onde um modelo se ajusta demasiadamente aos dados de treinamento. Isso significa que o modelo não apenas aprende os padrões subjacentes aos dados, mas também incorpora o ruído e as peculiaridades específicas do conjunto de treinamento.

Os principais hiperparâmetros utilizados para um bom treinamento do *Random Forest* são:

2.4.1.1 Número de Árvores (*n estimators*)

Este é o número de árvores de decisão que comporão o *ensemble*, que são o conjunto de modelos compostos por cada árvore que será utilizada. Um número maior de árvores geralmente leva a um modelo mais robusto, mas também pode aumentar o tempo de treinamento. No entanto, há um ponto de diminuição dos retornos, e adicionar muitas árvores além desse ponto pode não trazer benefícios significativos.

2.4.1.2 Profundidade Máxima das Árvores (*max depth*)

Controla a profundidade máxima de cada árvore de decisão, o que define quantos nós de decisão serão criados. Limitar a profundidade pode ajudar a evitar *overfitting*, especialmente em conjuntos de dados pequenos ou ruidosos.

2.4.1.3 Número Mínimo de Amostras em Folhas (*min samples leaf*)

Especifica o número mínimo de amostras necessárias em uma folha da árvore. Aumentar esse valor pode levar a árvores mais simples, reduzindo o *overfitting*.

2.4.1.4 Número Mínimo de Amostras para Divisão (*min samples split*)

Define o número mínimo de amostras necessárias para uma divisão em um nó interno da árvore. Isso pode influenciar a complexidade da árvore e também ajudar a evitar *overfitting*.

2.4.1.5 Número de Recursos a Considerar em Cada Divisão (*max features*)

Determina o número máximo de recursos a serem considerados em cada divisão de nó. Controlar esse número pode ajudar a reduzir a correlação entre as árvores e melhorar a diversidade do conjunto.

2.5 Redes neurais

As redes neurais, também conhecidas como redes neurais artificiais, são uma classe de modelos de aprendizado de máquina inspirados na estrutura e funcionamento do cérebro humano. Elas têm se destacado nas últimas décadas como uma ferramenta poderosa para resolver uma ampla gama de problemas em diversas áreas, como visão computacional, processamento de linguagem natural, reconhecimento de padrões, entre outras.

Como explicado por [Goodfellow, Bengio e Courville \(2016\)](#), o conceito de rede neural é baseado na ideia de que múltiplas unidades computacionais simples, chamadas de neurônios artificiais ou perceptrons, podem ser interconectadas em camadas para formar um sistema altamente adaptativo e capaz de aprender a partir dos dados. Cada conexão entre neurônios é associada a um peso, que determina a importância da entrada de um neurônio na saída de outro. Durante o treinamento, esses pesos são ajustados para otimizar o desempenho da rede em uma tarefa específica.

Uma das características mais notáveis das redes neurais é sua capacidade de realizar aprendizado profundo (*deep learning*), que envolve o uso de múltiplas camadas ocultas entre a camada de entrada e a camada de saída da rede. Essas camadas ocultas permitem que a rede aprenda representações hierárquicas complexas dos dados, o que é especialmente útil para tarefas de processamento de informações complexas, como reconhecimento de imagens ou tradução automática.

2.5.1 Hiperparâmetros de ajuste de Redes Neurais

Ao explicar RNAs, é importante destacar sua capacidade de aprendizado de padrões complexos e sua aplicabilidade em uma ampla gama de tarefas. Ajustar a arquitetura da RNA e seus hiperparâmetros é crucial para obter desempenho otimizado em diferentes problemas. Abaixo são demonstrados os principais hiperparâmetros utilizados em RNAs.

2.5.1.1 Número de Camadas e Neurônios por Camada

O número de camadas e neurônios em cada camada afeta a capacidade da rede de aprender padrões complexos, e é o que determina a arquitetura da rede. Uma arquitetura mais profunda com mais neurônios pode ser capaz de capturar representações mais abstratas, mas também pode aumentar o risco de *overfitting*.

2.5.1.2 Função de Ativação

Define a função utilizada pelos neurônios para introduzir não linearidades. Opções comuns incluem a função sigmoide, tangente hiperbólica (*tanh*) e unidades lineares retificadas (*ReLU*). A escolha depende do problema específico e pode impactar o treinamento e a capacidade de generalização. Neste trabalho é utilizada a função *ReLU*, recomendação

padrão de Goodfellow, Bengio e Courville (2016), que foi mantida após ter sido testada uma variação chamada *softplus* também recomendada, mas que obteve resultado pior.

2.5.1.3 Algoritmo de Otimização

Define o método usado para atualizar os pesos da rede durante o treinamento. Algoritmos comuns incluem o Gradiente Descendente, o Gradiente Descendente Estocástico (SGD), o Adaptive Moment Estimation (*Adam*) e o *RMSprop*. Cada algoritmo tem suas vantagens e desvantagens, e a escolha pode depender do problema. Como utilização padrão para vários tipos de modelos de *machine learning* devido sua robustez e adaptação, escolheu-se o *Adam*.

2.5.1.4 Batch Size

Define o número de amostras de treinamento utilizadas em uma iteração do algoritmo de otimização. Pode afetar a estabilidade do treinamento e a eficiência computacional. *Batch size* pequenos introduzem mais variação, enquanto *batch size* grandes podem requerer mais memória.

2.5.1.5 Épocas

Representa o número de vezes que todo o conjunto de treinamento é apresentado à rede. O treinamento ocorre em épocas, e ajustar esse parâmetro pode impactar o treinamento.

2.6 Métricas de desempenho de *machine learning*

As métricas de desempenho em *machine learning* são medidas quantitativas utilizadas para avaliar o quão bem um modelo de aprendizado de máquina está performando em uma tarefa específica. Essas métricas fornecem uma maneira objetiva de entender a eficácia do modelo, seja em tarefas de classificação, regressão ou outras. Como neste projeto se está trabalhando com previsão de dados financeiros, onde a tarefa é estimar um valor numérico, adotou-se dois principais estudos como base para selecionar as métricas analisadas: A monografia de Zaghi (2023), onde o autor utiliza o *machine learning* para prever preços de imóveis, e o artigo de Almeida Machado e Corrêa (2022) que avalia estas métricas para previsão de ativos financeiros, mas utilizando modelos diferentes. As métricas selecionadas são detalhadas a seguir.

2.6.1 Erro Absoluto Médio (MAE)

Calcula a média das diferenças absolutas entre as previsões e os valores reais. MAE é fácil de interpretar, pois representa diretamente a magnitude média do erro. No contexto

de previsão de receitas e despesas, o MAE fornecerá uma medida da precisão média do modelo em termos do valor absoluto das discrepâncias entre as previsões e os valores reais. Uma pontuação de MAE mais baixa indica melhor desempenho, pois menores diferenças absolutas sugerem que as previsões estão mais próximas dos valores reais. Seu cálculo pode ser feito de acordo com a equação a seguir:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.1)$$

Onde:

n é o número total de observações.

y_i é o valor observado na posição i .

\hat{y}_i é o valor previsto na posição i .

2.6.2 Erro Quadrático Médio (MSE)

Calcula a média das diferenças quadráticas entre as previsões e os valores reais. O MSE penaliza mais fortemente os erros grandes em comparação com o MAE. No contexto do projeto, o MSE avalia a precisão do modelo dando mais peso a discrepâncias maiores entre as previsões e os valores reais. Uma pontuação de MSE mais baixa é desejada, indicando que o modelo minimiza efetivamente os erros quadráticos e, portanto, está mais próximo dos valores reais. Seu cálculo pode ser feito de acordo com a equação a seguir:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.2)$$

Onde:

n = Número total de observações

y_i = Valor real na posição i

\hat{y}_i = Valor previsto na posição i

2.6.3 Coeficiente de Determinação (R^2)

Indica a proporção da variância na variável dependente que é explicada pela variável independente. Em contextos financeiros, R^2 pode ser útil para entender quanto da variabilidade nas previsões é capturada pelo modelo. Ele varia de 0 a 1, onde 1 indica que o modelo explica perfeitamente a variabilidade dos dados, e 0 indica que o modelo não oferece nenhuma melhoria em relação a uma simples média dos valores reais. Interpretar o R^2 ajuda a entender quanto da variação nas receitas e despesas é explicada pelos modelos. Um R^2 mais alto sugere um ajuste melhor do modelo aos dados. Seu cálculo pode ser feito de acordo com a equação a seguir:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (2.3)$$

Onde:

SS_{res} = Soma dos quadrados dos resíduos

SS_{tot} = Soma total dos quadrados

3 Desenvolvimento

3.1 Planejamento da aplicação

Iniciando o desenvolvimento da aplicação, definiu-se os dados que serão requisitados para o usuário, para definir assim as tabelas e itens da base de dados. Criou-se então o seguinte escopo de tabelas:

- Usuários: Tabela contendo demais nomes em que o usuário pode atribuir as compras e demais transações para este nome específico, função vai auxiliar na divisão de despesas para aqueles que organizam finanças de mais de uma pessoa, por exemplo o chefe da família;
- Cartões: Tabela contendo dados necessários dos cartões do usuário, para atribuir a compra a um cartão específico, e utilizar as datas de fechamento e vencimento para gerenciar os dados mensais do usuário;
- Categorias: Tabela contendo a lista de categorias de compras que cada usuário irá definir, para auxiliar na categorização das transações;
- Compras de cartão: Tabela contendo compras realizadas nos cartões do usuário;
- Transações: Tabela contendo demais despesas e receitas fora dos cartões;

Definido o escopo da base de dados, estas tabelas foram definidas no *Microsoft Excel*.

Para auxiliar no acesso às informações da base de dados, as foram criadas tabelas dinâmicas a partir das tabelas *Compras de cartão* e *Transações*. Esta ferramenta nos permite a organização dos dados de uma planilha promovendo a transformação de dados em informações objetivas disponíveis em relatórios com conteúdos dinâmicos sem a necessidade de criar funções complexas ([TRYBE, s.d.](#)).

3.2 *Dashboard excel*

A aplicação no Excel possui 1 planilha para inserção de dados, chamada Cadastro de dados. Nela tem-se as tabelas *Lista de cartões*, onde o usuário cadastra o nome, data de fechamento da fatura e vencimento de cada um de seus cartões, a tabela *Usuários*, onde podem ser cadastrados mais nomes de pessoas para vincular as receitas e despesas, e a tabela *Categorias*, onde são cadastradas as categorias de compras que o usuário irá utilizar.

Valor total	Data	Parcelas	Cartão	Usuário	Categoria	Descrição	Valor da Parcela	Fechamento da compra	Data primeira parcela	Parcelas abertas	Data última parcela
RS 100,00	17/01/2023	1	Bradesco	João	saude	exame vista	RS 100,00	0	17/01/2023	0	28/02/2023

Figura 1 – Trecho de fórmula da base de dados

Fonte: Autoria própria.

Estas tabelas serão utilizadas para preencher a base de dados, sendo assim, o usuário só poderá associar uma despesa ou receita a um cartão, usuário e categoria cadastrados anteriormente.

A partir das tabelas citadas anteriormente, são feitas algumas verificações adicionais a partir dos dados cadastrados para cada transação. Como exemplo, para as transações vindas de cartão de crédito é feito o cálculo de qual é a fatura que esta compra será lançada, pois caso sua data seja a partir da data de fechamento, será lançada na fatura seguinte. Outra análise que é feita, é data da última fatura, útil para auxiliar na análise de compras parceladas. Estes dados são salvos direto na base de dados, em colunas adicionais atualizadas conforme a inserção de novas compras. A Figura 1 demonstra um trecho da base de dados, onde a coluna em vermelho com a fórmula destacada descreve se o lançamento da compra iniciará no mês corrente ou mês seguinte.

Além disso adicionou-se 3 principais planilhas para visualização sendo elas *Lançamentos cartões*, *Demais lançamentos* e *Visão simplificada*.

A planilha *Lançamentos cartões* contém as informações detalhadas das compras de cartões, separando as compras pelo mês escolhido, de acordo com a data da compra e fechamento da fatura. Este filtro garante que toda as compras realizadas no período entrem nesta fatura, coincidindo o valor real que vem nas fatura de cartão de crédito.

A planilha *Demais lançamentos* contém as informações detalhadas de despesas e receitas fora dos cartões. Ela conta com o mesmo filtro utilizado anteriormente, então todas despesas e receitas realizadas no período entram na visualização.

A planilha *Visão simplificada* conta com gráficos interativos de despesas e receitas gerais, sem detalhamento de lançamentos. Ela fornece ao usuário uma visão simplificada das despesas e receitas, além de separar os gastos por categoria.

3.3 Manipulação de dados para os modelos de previsão

A base de dados foi adquirida de um projeto autoral para gestão de finanças, onde tem-se duas bases de dados principais, uma de compras de cartões de crédito, e outra de demais despesas e receitas. Estas bases foram unificadas a partir da aplicação em Excel, para gerar uma nova base de dados única para o treinamento dos modelos. A base de dados unificada possui 518 amostras, separadas em linhas, e contém informações de um período de 7 meses.

Utilizou-se a plataforma de desenvolvimento *Google Colaboratory* (GOOGLE, 2024), que permite a execução de códigos na linguagem *Python*. Para isto a planilha vinda do Excel foi salva no *Google Drive*, e importada no ambiente de desenvolvimento, para que por meio da biblioteca *Pandas* (PANDAS, 2024) do *Python*, fosse possível converter a planilha em um *DataFrame*, para ser possível manipulá-la para preparar a base para o treinamento do modelo. Como a base de dados não era preparada para este tipo de utilização, grande parte dos dados estavam descategorizados, e como nossa meta é fazer a previsão do valor das compras por categoria, foi necessária a categorização manual de grande parte dos dados, tarefa repetitiva mas crucial para que faça sentido utilizar os modelos de *Machine Learning* para prever corretamente os valores por categoria.

3.4 Treinamento do modelo *Random Forest*

O intuito da utilização dos modelos de *machine learning* foi utilizá-los para prever o valor de lançamentos na nossa base de dados. Assim tem-se o atributo alvo do modelo definido, o preço. Já para os dados de treinamento, foram definidas as *features* do modelo, que serão os atributos utilizados para o treinamento do modelo, tendo como alvo o preço do lançamento. Para estas *features* foram selecionados os dados de data, classificação se é uma despesa ou receita, número da parcela do lançamento e categoria do lançamento.

O modelo *Random Forest* foi implementado por meio da biblioteca *scikit-learn* (SCIKIT-LEARN, 2024), também utilizando o *Google Colaboratory* (GOOGLE, 2024). Os hiperparâmetros do modelo foram ajustados após testes analisando as métricas de desempenho MAE, MSE e R^2 . Os dados de treino e teste foram divididos respectivamente em 70% e 30% da base de dados, e no melhor resultado testado obteve-se o *N Estimators* definido como 300, Max depth como 100, Min Samples Leaf como 1, e *Min Samples Split* como 25. Estes e os demais hiperparâmetros utilizados estão descritos na Tabela 1:

Hiperparâmetro	Valor
Features	despesaOuReceita (indica se a amostra é uma despesa ou receita), dia da semana, mês, categoria (indica categoria da amostra)
Target	preço (preço da amostra em reais)
Treinamento/Teste	70% treinamento, 30% teste
Imputer	SimpleImputer(strategy='constant', fill_value=1)
Número de Estimadores	300
Profundidade Máxima	100
Folhas Mínimas por Folha	1
Divisões Mínimas por Nó	25

Tabela 1 – Hiperparâmetros do Modelo *Random Forest*.

3.5 Treinamento do modelo baseado em Rede Neural

Para o modelo de rede neural, assim como no de floresta aleatória foram mantidos o atributo alvo como o preço, e as *features* do modelo como dados de data, classificação se é uma despesa ou receita, número da parcela do lançamento, e categoria do lançamento. O modelo de rede neural foi implementado utilizando a biblioteca *Keras* (KERAS, 2024), utilizando a linguagem de programação *Python*. Assim como no modelo baseado em *Random Forest*, os dados de treino e teste foram divididos respectivamente em 70% e 30% da base de dados. De acordo com a complexidade da base de dados é necessária a utilização de mais camadas para a rede neural, este número foi definido após testes em que 6 camadas mostrou maior assertividade, uma vez que o aumento de camadas para este treinamento aumentou a função de perda do algoritmo. O número de neurônios foi definido de acordo com padrões normalmente utilizados, em que a camada central tende a ter maior número de neurônios, e suas adjacentes reduzem pela metade o número de neurônios, até que na última camada temos somente 1 neurônio, para que seja selecionada a saída com melhor taxa de aprendizado, como mostrado na Tabela 2. A função de ativação utilizada em cada camada foi a *ReLU* e o algoritmo de otimização utilizado foi o *Adam*, buscando minimizar o erro médio quadrático, geralmente utilizado em modelos de regressão.

Hiperparâmetro	Valor
Features	despesaOuReceita (indica se a amostra é uma despesa ou receita), dia da semana, mês, categoria (indica categoria da amostra)
Target	preço (preço da amostra em reais)
Treinamento/Teste	70% treinamento, 30% teste
MinMaxScaler	MinMaxScaler()
Imputer strategy	SimpleImputer(strategy='constant', fill_value=1)
Número de Camadas	6
Neurônios por Camada	32, 64, 128, 64, 32, 1
Função de Ativação	relu
Número de Épocas	200
Batch size	14
Função otimização	adam
Função de Perda	mean_squared_error

Tabela 2 – Hiperparâmetros da Rede Neural.

3.6 Fluxo de dados

O fluxo de funcionamento do projeto por ser visualizado na Figura 2. A partir das tabelas cartões, categorias e usuários, os dados são inseridos nas bases de dados de crédito e demais despesas e receitas. No Excel são realizados os principais cálculos para utilizar

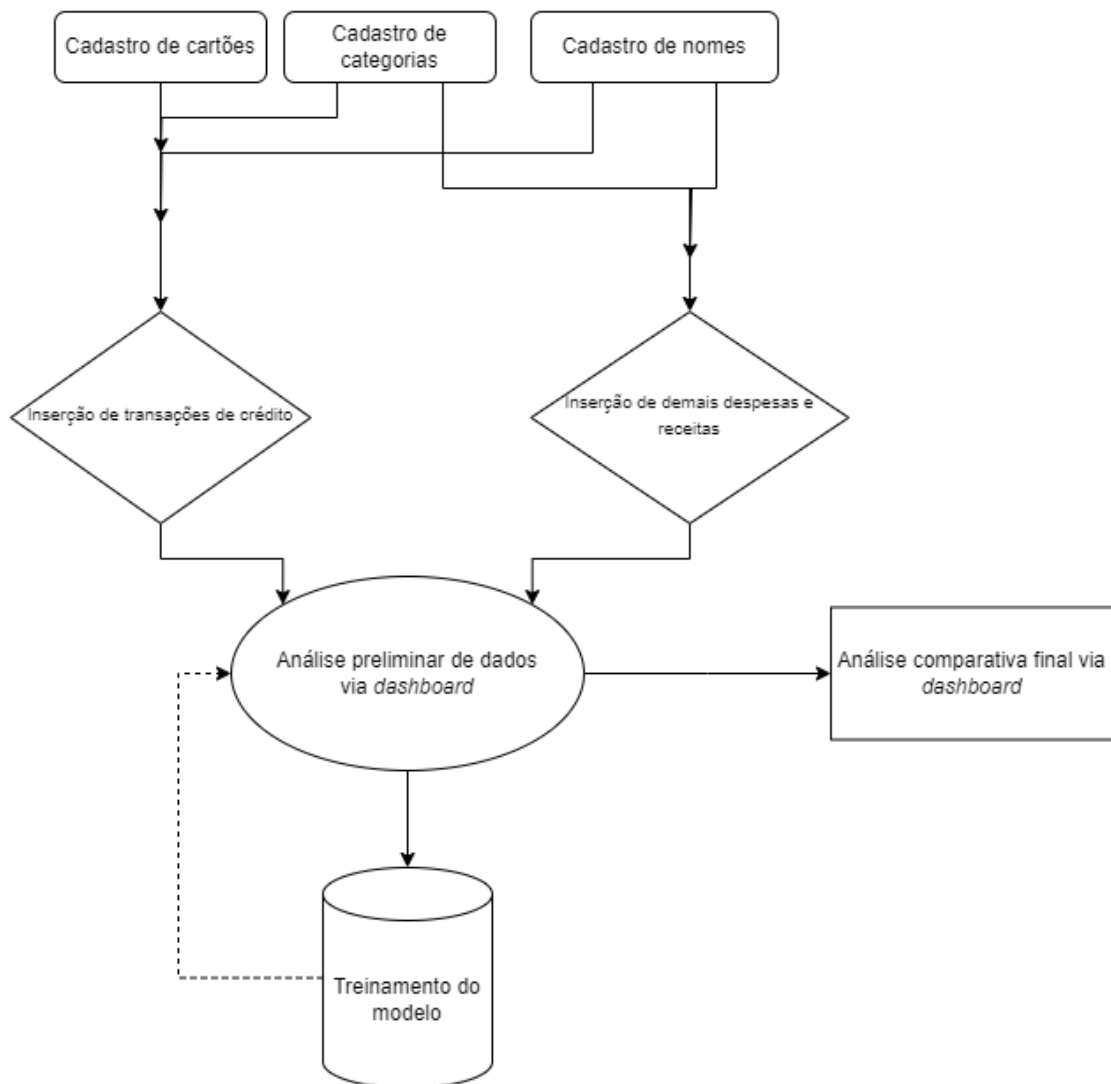


Figura 2 – Fluxograma de processos do projeto

Fonte: Autoria própria.

o filtro das transações a partir do mês, usuário e cartões, além de análise categorizada e gráfico de despesas ao longo dos meses. A base de dados é então importada e tratada para treinamento dos modelos de *machine learning*. Escolhido o modelo com melhor desempenho, suas previsões são importadas no *dashboard* para mostrar uma previsão de gastos para o usuário.

4 Resultados

4.1 Planilhas de visualização

As planilhas foram confeccionadas por filtros e tabelas dinâmicas a partir da base de dados do Excel. Na planilha *Cadastro de dados* utilizou-se o VBA para confecção de botões para salvar os lançamentos na base de dados, como mostrado na Figura 3, e também definiu-se as tabelas *Lista Cartões*, *Usuários* e *Categorias* como mostrado na Figura 4.

Cadastrar compra cartão	
Valor total	
Data	
Parcelas	
Cartão	
Usuário	
Categoria	
Descrição	
<div style="display: flex; justify-content: space-around;"> GERAR SALVAR </div>	

Figura 3 – Trecho onde os dados do usuário são inseridos para serem salvos na base de dados.

Fonte: Autoria própria.

Lista cartões		
Cartão	Dia fechamento	Vencimento
Bradesco	31	10
Santander	9	15

Usuários
Lucas
João

Categorias
Sem categoria
Recorrente
Alimentação
Internet
sacolão
saude

Figura 4 – Tabelas base para preenchimento das transações financeiras no Excel.

Fonte: Autoria própria.

Fatura do mês		fev-23		Cartão: Bradesco				Total:	R\$ 2.390,28
Filtro									
Valor total	Data	Nº Parcelas	Cartão	Usuário	Categoria	Descrição	Valor da Parcela	Parcela	
R\$ 175,42	20/12/2022	2	Bradesco	João	Cachorro		R\$ 87,71	2 de 2	
R\$ 108,66	20/12/2022	2	Bradesco	João	Cachorro		R\$ 54,33	2 de 2	
R\$ 87,44	20/12/2022	2	Bradesco	João	Roupa		R\$ 43,72	2 de 2	
R\$ 270,00	20/12/2022	3	Bradesco	João	Academia		R\$ 90,00	2 de 3	
R\$ 60,81	20/12/2022	3	Bradesco	João	casa		R\$ 20,27	2 de 3	
R\$ 124,86	20/12/2022	3	Bradesco	João	casa		R\$ 41,62	2 de 3	
R\$ 106,20	01/01/2023	1	Bradesco	João	sem categoria		R\$ 106,20	1 de 1	
R\$ 50,04	20/01/2023	1	Bradesco	João	sem categoria		R\$ 50,04	1 de 1	

Figura 5 – Trecho da planilha com busca de lançamentos do cartão

Fonte: Autoria própria.

Selecione os dados:	
	Fatura
Ano:	2023
Mês:	2
Cartão:	Bradesco
Usuário:	João

Figura 6 – Seletor de dados da planilha de lançamento de cartões.

Fonte: Autoria própria.

Na planilha *Lançamentos Cartões*, mostrada na Figura 5. Este filtro busca a partir de um seletor de dados, Figura 6, os lançamentos por cartão, por usuário e por mês, mostrando uma busca mais detalhada de todas as transações do período. O seletor utiliza a ferramenta de validação de dados do Excel, então, as opções que se podem selecionar nos filtros são as pré cadastradas nas tabelas de dados como comentado no item 3.2. É possível também realizar o filtro de todos os cartões unificados, caso se tenha mais de um. A mesma lógica é aplicada ao filtro de busca de lançamentos fora do cartão de crédito.

Por fim, na planilha *Visão simplificada* tem-se os valores da faturas e demais lançamentos para a projeção dos próximos 12 meses, além da visualização do montante de compras por categoria. Estas informações são ainda convertidas em gráficos, como mostrado na Figura 7.

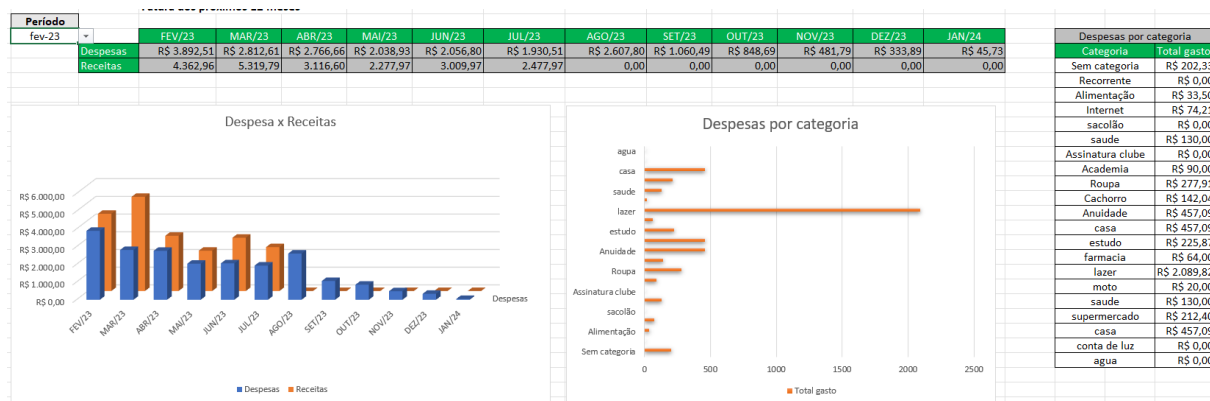


Figura 7 – Planilha que reúne a visão simplificada dos dados do usuário.

Fonte: Autoria própria.

4.2 Análise dos resultados do modelo de *Random Forest*

Após o treinamento do modelo, foram analisados os resultados das métricas utilizadas para a validação dos modelos. Para essa análise, foram verificados os resultados das métricas erro médio absoluto (MAE), erro quadrático médio (MSE) e coeficiente de determinação R^2 . Os resultados foram: MAE de 69.83 Reais, MSE de 21865,26 Reais² e R^2 de 0,70, como mostrado na Tabela 3.

Tabela 3 – Resultados do Modelo *Random Forest*

Métrica	Valor
Mean Absolute Error (MAE)	R\$ 69.83
Mean Squared Error (MSE)	R\$ ² 21865.26
R-squared (R^2)	0.70

Fonte: Autoria própria.

4.3 Análise dos resultados do modelo de Redes Neurais

Assim como citado para o modelo de *Random Forest*, para o modelo de Redes Neurais foram analisados os resultados das métricas utilizadas para a validação dos modelos utilizados. Para essa análise, foram utilizadas as métricas já descritas anteriormente, são elas: erro médio absoluto (MAE), erro quadrático médio (MSE) e coeficiente de determinação (R^2). Os resultados foram o MAE de 86,20 Reais, o MSE de 28920,08 Reais² e o R^2 foi de 0,60, como mostrado na Tabela 4.

Tabela 4 – Resultados do Modelo de Rede neural

Métrica	Valor
Mean Absolute Error (MAE)	R\$ 86.20
Mean Squared Error (MSE)	R\$ ² 28920.08
R-squared (R ²)	0.60

Fonte: Autoria própria.

4.4 Comparativo entre os modelos

Após a análise dos dois modelos, tem-se o *Random forest* como modelo que melhor previu os valores dos lançamentos, tendo um coeficiente de determinação R² significativamente maior que do modelo de redes neurais, comparação é mostrada na Tabela 5.

Tabela 5 – Comparação de Resultados entre Random Forest e Rede Neural

Métrica	Random Forest	Rede Neural
Mean Absolute Error (MAE)	R\$ 69.83	R\$ 86.20
Mean Squared Error (MSE)	R\$ ² 21865.26	R\$ ² 28920.08
R-squared (R ²)	0.70	0.60

Fonte: Autoria própria.

4.5 Problemas ao utilizar os dados previstos

Selecionado o *Random forest* como modelo que melhor prevê os dados, foi observado um problema. Para que os dados previstos por ele sejam utilizados no *dashboard*, as manipulações dos dados que foram feitas devem ser desfeitas, para que os dados correspondam com os dados da planilha original. Alguns exemplos são a coluna *Categorias*, em que foi utilizada a técnica chamada *one-hot encoding*, que transforma cada tipo de categoria em uma coluna nova do *dataframe*, e a correspondência deste lançamento na categoria, é indicado por um valor binário. O mesmo ocorre para a data que foi dividida em mês e dia da semana, para que o modelo possa tentar encontrar padrões. Essas manipulações foram necessárias pois como a base de dados é extremamente limitada, sem estas transformações de dados o aprendizado do modelo não obteve bons resultados. Após análise das previsões, foi possível notar também que para algumas categorias não foi gerada nenhuma previsão, o que atrapalharia sua utilização.

Uma opção seria unir todos os dados por categoria e mês, para que seja previsto um único valor por categoria e mês de cada correspondência, assim seria possível prever um montante para o mês de cada categoria, mas esta opção não foi viável, pois não há

dados suficientes para treinamento. Algumas tentativas sem sucesso, fizeram com que esta retransformação dos dados fosse uma barreira para utilizar o modelo como previsão no *dashboard*.

5 Conclusões

A proposta deste trabalho foi criar uma aplicação para auxiliar na gestão e controle de finanças dos usuários, com o auxílio de ferramentas do *Microsoft office*, e realizar um estudo entre modelos de inteligência artificial, para identificar qual modelo prevê melhor os lançamentos financeiros do usuário, para que auxilie o usuário prevendo suas finanças.

Entre os objetivos descritos em 1.2.2, foi possível concluir o *dashboard* para análise financeira. Além disso, realizou-se o treino e comparação entre os modelos *Random Forest* e redes neurais, sendo o *Random Forest* o que apresentou melhor assertividade que o modelo de redes neurais, ficando, significativamente maior do que o modelo de redes neurais, ficando para novas abordagens deste trabalho a integração das previsões do modelo juntamente ao *dashboard*.

Vale ressaltar que o *dashboard* mostrou-se bastante válido para auxiliar no controle das finanças, uma vez que os dados inseridos são utilizados para gerar montantes de despesas e receitas, mostrar graficamente os gastos em cada mês, além de mostrar gastos por categorias, auxiliando o usuário a identificar possíveis desperdícios ou gastos que devem ser analisados com cautela por estar fugindo dos padrões.

5.1 Trabalhos futuros

Como recomendações de trabalhos futuros, pode-se citar a integração das previsões vindas do modelo de aprendizado de máquina, tarefa que não foi implementada neste trabalho. Esta integração pode ser feita de modo que periodicamente, ao ser alimentada com novos dados, o modelo seja novamente treinado, e assim sejam geradas novas previsões. A tentativa de outro método de ajuste da base de dados para o treinamento também é recomendada, em específico, pode-se citar a alteração do parâmetro *shuffle* na implementação do modelo *Random Forest* no *scikit-learn*. Este parâmetro é usado para determinar se as amostras fornecidas ao estimador são embaralhadas antes de cada árvore ser ajustada. Seu valor padrão é *True*, o que não foi alterado neste trabalho, mas pode gerar resultados interessantes para o treino no modelo.

Pode-se também implementar um modelo baseado em Séries Temporais, abordagem que pode ser interessante, uma vez que nossas previsões utilizam também a informação de data para previsão do modelo. Outra opção recomendada é utilizar uma base de dados maior, para que não seja necessário utilizar tantas *features* para o treinamento dos modelos, o que pode facilitar a integração das previsões futuramente. Outro possível teste seria realizar o treinamento com uma base de dados para cada categoria, garantindo que para

cada categoria haverá pelo menos uma correspondência naquele mês, e a previsão seja uniforme entre as categorias. É interessante também o teste de outros modelos de *machine learning* ou métodos de validação são também válidos, para analisar diferentes abordagens.

Por último, outra sugestão é buscar integrar o *dashboard* às informações bancárias do usuário, de forma a se ter o preenchimento automático dos dados. Por exemplo, o usuário fornecer a fatura ou extrato bancário e a aplicação identificar automaticamente os dados e preenche as tabelas. Esta funcionalidade pode ser observada como tendência nos aplicativos de bancos, onde com o consentimento do usuário em compartilhar seus dados pelo *Open Finance*, o aplicativo reúne as informações dos demais bancos em um só aplicativo, facilitando a visualização e gestão do usuário.

Referências

ALMEIDA MACHADO, André Augusto de; CORRÊA, Geraldo Nunes. *Avaliação de modelos de previsão dos valores das ações no mercado financeiro usando aprendizado de máquina*. RETEC. 2022. Disponível em: <https://www.fatecourinhos.edu.br/retec/index.php/retec/article/view/440>. Acesso em: 10 dez. 2023. Citado 1 vez na página 20.

AVELAR, Ewerton Alex et al. Algoritmo Random Forest para Previsão de Comportamento de Preços de Ativos. *revistafsa*, p. 21, 2022. Disponível em: <http://www4.unifsa.com.br/revista/index.php/fsa/article/view/2592/491493498>. Citado 1 vez na página 17.

BALTZAN, Paige; PHILLIPS, Amy. *Sistemas de informação*. AMGH, 2012. Citado 3 vezes na página 16.

BISHOP, Christopher M. *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006. Citado 1 vez na página 16.

G1, Portal. *Endividamento e inadimplência das famílias batem novo recorde em abril*. Portal G1. 2022. Disponível em: <https://g1.globo.com/economia/noticia/2022/05/02/endividamento-e-inadimplencia-das-familias-batem-novo-recorde-em-abril.ghtml>. Acesso em: 17 mar. 2023. Citado 1 vez na página 13.

GHOSH, Indranil; JANA, Rabin K; SANYAL, Manas K. Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using nonlinear dynamics, econometric models and machine learning algorithms. *Applied Soft Computing*, Elsevier, v. 82, p. 105553, 2019. Citado 1 vez na página 17.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org/>. Citado 2 vezes nas páginas 19, 20.

GOOGLE. *Colaboratory Documentation*. 2024. [S.l.: s.n.] https://colab.research.google.com/notebooks/wBRscrollTo=5fCEDCU_qrC0. Citado 2 vezes na página 25.

KERAS. *Keras Documentation*. 2024. [S.l.: s.n.] <https://keras.io/api/>. Citado 1 vez na página 26.

MACEDO, Jurandir Sell. *A árvore do dinheiro: guia para cultivar sua independência financeira*. Elsevier, 2007. Citado 1 vez na página 12.

PANDAS. *Pandas Documentation*. 2024. [S.l.: s.n.] <https://pandas.pydata.org/docs/>. Citado 1 vez na página 25.

- QUINTINO, Joyce et al. Avaliação de Desempenho de Algoritmos de Machine Learning para Otimização de Simulações de Redes de Computadores. In: ANAIS do XXV Workshop de Gerência e Operação de Redes e Serviços. Rio de Janeiro: SBC, 2020. P. 167–180. DOI: [10.5753/wgrs.2020.12459](https://doi.org/10.5753/wgrs.2020.12459). Disponível em: <https://sol.sbc.org.br/index.php/wgrs/article/view/12459>. Citado 1 vez na página 17.
- ROBBINS, Stephen Paul. *Administração: mudanças e perspectivas*. Saraiva, 2000. Citado 2 vezes nas páginas 12, 16.
- SCIKIT-LEARN. *scikit-learn Doc*. 2024. [S.l.: s.n.] <https://scikit-learn.org/0.21/index.html>. Citado 1 vez na página 25.
- STATUSINVEST. [S.l.: s.n.] <https://statusinvest.com.br/>. Citado 1 vez na página 16.
- TORO. *mobills para gestão financeira*. [S.l.: s.n.] <https://www.mobills.com.br/sobre/>. Citado 1 vez na página 16.
- TRYBE. Tabela dinâmica: o que é e como fazer no Excel? Trybe. Citado 1 vez na página 23.
- TURBAN, Efrain; MCLEAN, Ephraim; WETHERBE, James. *Tecnologia da informação para gestão*. 3. ed.: Bookman, 2004. Citado 1 vez na página 12.
- XPEDUCAÇÃO. *Importância do controle financeiro: o que é e como adotar esse hábito sem complicação?* XP Educação. 2021. Disponível em: <https://blog.xpeducacao.com.br/importancia-do-controle-financeiro/>. Acesso em: 17 mar. 2023. Citado 1 vez na página 12.
- YOSHINAGA, Claudia Emiko; CASTRO, F Henrique. Inteligência artificial: a vanguarda das finanças. *GV-EXECUTIVO*, v. 22, n. 3, 2023. Citado 1 vez na página 12.
- ZAGHI, Lucca Magri. *Modelo de previsão de preços de imóveis na cidade de Florianópolis/SC a partir de técnicas de Machine Learning*. UFSC. 2023. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/248689/TCC.pdf?sequence=4&isAllowed=y>. Acesso em: 10 jan. 2024. Citado 1 vez na página 20.
- ZULIAN, Maurício. *EXCEL/VBA para gerenciamento de micro e pequenas empresas*. 2013. Disponível em: <https://ojs.fho.edu.br:8481/revfho/article/view/118/114>. Acesso em: 17 out. 2023. Citado 1 vez na página 12.