**Projeto Aplicado – Relatório Final**

|  |  |
| --- | --- |
| **Nome do Aluno** | **ADRIANO MARABUCO DE ALBUQUERQUE LIMA** |
| **Título do Trabalho** | **GESTÃO DE INVESTIMENTOS ORIENTADA POR APRENDIZADO DE MÁQUINA** |
| **Curso** | MBA em Aprendizado de Máquina |
| **Linha de Especialização** | Machine Learning Aplicado aos Sistemas de Recomendação |
| **Orientador** | **BRUNO RAFAEL DE OLIVEIRA RODRIGUES** |
| **Data** |  |

**INTRODUÇÃO**

1. Apresentação do Desafio e da Solução:
   1. Setor do mercado e a justificativa de tal seleção;
   2. Características e restrições de escopo do Desafio;
   3. Oportunidade vislumbrada que motivou o desenvolvimento da Solução.

a. Setor do mercado e a justificativa de tal seleção;

O setor de gestão de investimentos dedica-se a alocação de capital em ativos em busca de remuneração pela disponibilização desses recursos. Ele desenvolveu-se bastante no Brasil, nos últimos anos, com a estabilização econômica e evolução do mercado de capitais.

É um setor muito competitivo e globalizado que movimenta grandes fortunas diariamente entre diferentes ativos e mercados, composto por instituições financeiras, fundos de investimento e particulares.

Divide-se basicamente em dois tipos de investimento: renda fixa e renda variável. O mercado renda fixa é direcionado para investimentos remunerados a título de juros, como títulos públicos, debêntures e outros instrumentos de dívida. Enquanto o mercado de renda variável é direcionado para investimentos remunerados como participação nos lucros e ganho de capital, como ações, commodities e derivativos.

O objetivo da gestores de renda variável é obter retornos iguais a média do mercado (beta) ou acima da média do mercado (alfa). Os investimentos com retornos alfa elevados costumam atraem mais clientes, que procuram as alternativas mais rentáveis para suas aplicações financeiras.

O ambiente de investimento de renda variável marcado por alta volatibilidade e incerteza torna bastante desafiador obter retorno alfa significativo, e muitos acabam sofrendo grandes perdas na busca por esse resultado diferenciado. No entanto aqueles que obtém tais retornos podem usufruir de prestígio e amealhar fortunas.

Em 2007 e 2008 trabalhei em uma gestora de investimentos chamada TARPON, que é especializada em renda variável com uma abordagem ativista, em que busca influenciar no negócios em prol da maximização de valor. Atuei na área de análise de investimentos realizando avaliação de negócios, bem como monitoramento de performance e mercado.

Naquela época alguns fundos começaram a criar estratégia quantitativas de investimento que utilizam métodos estatísticos e programação para melhorar o resultado dos investimentos. Desde então essa tendência ganhou força, e cada vez mais o mercado de gestão de investimentos tem utilizado ferramentas de automação e inteligência artificial em suas atividades. Como é um setor bastante dinâmico e digitalizado, há uma abundância de dados a serem trabalhados, o que torna um campo fértil para projetos de machine learning.

A minha experiência profissional no setor aliada as condições favoráveis tornam o setor de gestão de investimentos um bom alvo para desenvolver meu projeto de machine learning.

b. Características e restrições de escopo do Desafio;

O desafio é criar um projeto de aprendizado de máquina capaz de prever os preços das ações, que é um fator que determina o sucesso ou fracasso de muitos investimentos, pois uma previsão errada leva o gestor a tomar decisões de investimentos baseada numa premissa equivocada e obter um resultado adverso.

Esse desafio está sujeito a dificuldades inerentes às incertezas e à volatidade do mercado. Há especialistas que afirmam ser impossível prever os movimentos do mercado, porém, estudos recentes efetuados por cientistas de dados revelam ser possível identificar padrões nos movimentos de mercado, gerando previsibilidade.

Como forma de reduzir a complexidade do projeto não serão realizadas previsões de derivativos que são ativos mais complexos de se trabalhar do que as ações. Assim o foco principal será a previsão do preço de ações. O planejamento das estratégias de investimento e formação de carteiras também estão fora do escopo, apesar de estarem relacionados na realização dos investimentos.

c. Oportunidade vislumbrada que motivou o desenvolvimento da Solução.

A gestão de investimentos lida com diversos desafios cujos impactos podem levar o negócio a ruína. Riquezas acumuladas durante muitos anos podem sumir em questão de dias. Os gestores de investimentos vivem na “corda bamba” entre ganhar fortunas e perdê-las pelos motivos que se seguem: alterações súbitas no comportamento de um ativo; exposição superestimada ou subestimada do risco; e falhas no momento de compra ou venda.

Os investimentos de renda variável são caracterizados pela volatilidade dos preços dos ativos, o que é resultado das variações de preços das compras e vendas. O problema é que essa variável também oscila com o tempo, portanto tem uma distribuição não estocástica, gerando grande incerteza nas alterações dos preços, pois eles podem sair de um patamar de baixa volatilidade para alta de um dia para o outro, e um investimento de baixo de risco tornar-se de alto. Essas alterações súbitas das variações de preços podem tornar um investimento rentável em um fracasso com grandes perdas financeiras.

Essa ruptura no padrão de negociação, e risco, gera outro problema na gestão dos investimentos, pois interfere no binômio risco/retorno do ativo, e de forma agregada no portfólio de ativos, criando uma exposição indesejada ao risco. Investidores agressivos buscam maiores retorno e estão dispostos a suportarem mais riscos, enquanto o oposto vale para os conservadores. Ao definir a estratégia de investimento, o gestor busca adequar os investimentos ao perfil do público-alvo dos investidores, entretanto mudanças no mercado podem provocar um desalinhamento disso. Por exemplo, o gestor pode comprar determinada ação buscando montar uma carteira de baixo risco e baixo retorno para investidores mais conservadores, e por um aumento repentino na volatilidade sofrer grandes perdas monetárias com a queda no preço dos ativos que se tornarem instáveis.

Outra questão que também deriva da volatilidade é o timing de compra e venda dos ativos. A fórmula para o sucesso no mercado financeiro é simples: comprar na baixa e vender na alta. O problema é que na prática é muito difícil acertar a hora de comprar e vender, e muitas vezes acontece o oposto. Após um alta contínua nos preços, muitos negociadores resolvem comprar o ativo achando que a alta vai persistir, e acaba acontecendo o contrário, ou após uma grande queda, muitos ficam receio de a queda continuar e se desfazem de suas ações, ficando com o prejuízo e perdendo a recuperação dos preços. Essas falhas de timming dilapidam o patrimônio dos fundos de investimento.

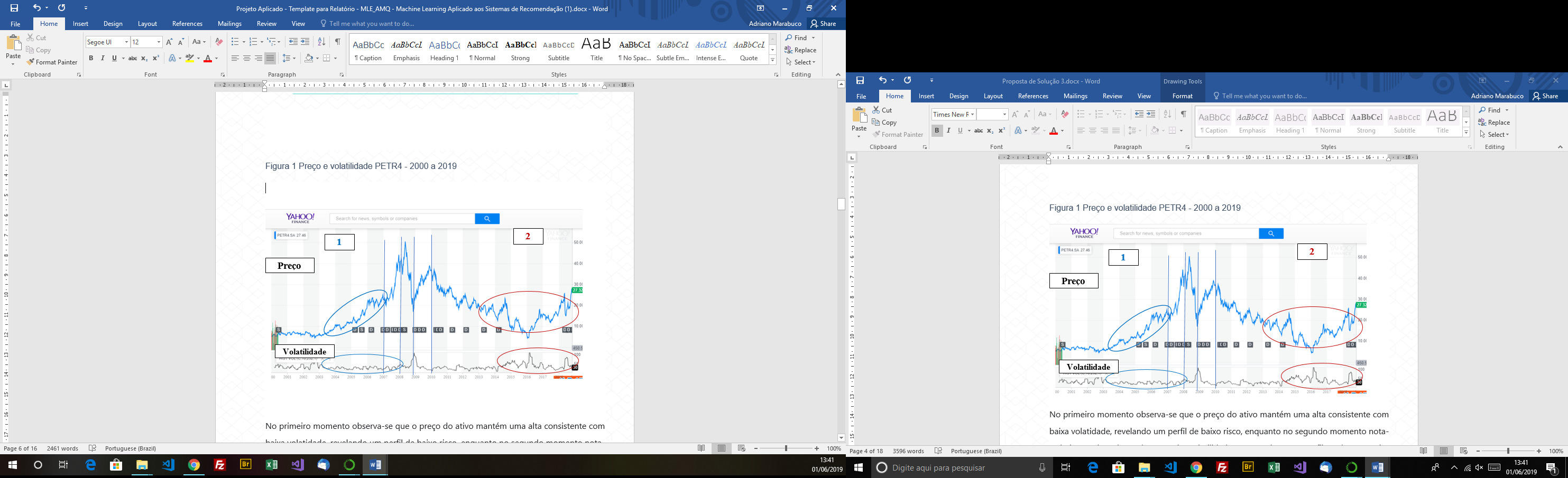
Os desafios elencados anteriormente são evidenciados por casos notáveis de crises em fundos de investimento, assim como pelas informações de mercado.

No que tange a variação do padrão de comportamento dos ativos e seus efeitos é ilustrativo o caso do LTCM, que foi um grande fundo de investimento americano idealizado e gerenciado por economistas ganhadores do prêmio do Nobel (Myron S. Scholes e Robert C. Merton), que usavam estratégias de investimento baseadas em modelos econométricos avançados, e foram surpreendidos por mudanças no comportamento do mercado que provocaram um caos financeiro e a bancarrota do fundo e muitos investidores:

*Initially successful with annualized return of over 21% (after fees) in its first year, 43% in the second year and 41% in the third year, in 1998 it lost $4.6 billion in less than four months following the 1997 Asian financial crisis and 1998 Russian financial crisis, requiring financial intervention by the Federal Reserve, with the fund liquidating and dissolving in early 2000.  
Fonte: https://en.wikipedia.org/wiki/Long-Term\_Capital\_Management.*

A alteração no binômio risco/retorno pelas mudanças no comportamento dos ativos pode ser identificada pela análise de alguns casos, como o da Petrobrás, conforme o gráfico a seguir.

Figura 1 Preço e volatilidade PETR4 - 2000 a 2019



No primeiro momento observa-se que o preço do ativo mantém uma alta consistente com baixa volatidade, revelando um perfil de baixo risco, enquanto no segundo momento nota-se baixas e altas alternadas com alta volatilidade, mostrando que o perfil mudou para alto risco, quando comparado ao primeiro momento. Como mencionado anteriormente, a relação risco/retorno do ativo mudou ao longo dos anos, então o gestor de investimentos que comprou a estável no Petrobrás nos anos de 2003 a 2007 foi surpreendido por um pico de volatilidade em 2009, e um aumento consistente na volatilidade a partir de 2014, tornando-o um ativo instável.

A evidência do timming também pode ser inferida do gráfico exposto. Quem comprou no final de 2007 e vendeu em meados de 2008, auferiu um bom lucro, porém quem esperou para vender em 2009 pode ter saído no prejuízo, e quem se empolgou com a alta dos preços e comprou em meados de 2008, depois se assustou com a queda e vendeu no início de 2009 ficou um grande prejuízo, mas se esperou para vender em 2010 reduziu sua perda.

Diante do exposto, percebe-se que as alterações súbitas no comportamento de um ativo; exposição superestimada ou subestimada do risco; e falhas no momento de compra ou venda são dores reais sofridas pelos investidores do mercado financeiro.

A utilização do aprendizado de máquina para identificar pelos dados históricos: os padrões de comportamento dos ativos, antecipar suas variações e prever preços pode oferecer uma forma de mitigar os riscos das decisões de investimentos e facilitar a vida dos gestores e investidores.

Existem empresas já trabalhando com esse tipo de solução, conforme relata a matéria do jornal valor econômico “Inteligência artificial facilita a vida de investidores” de 31/01/19:

“Seival Investimentos e Visia estão entre as gestoras de fundos baseados na tecnologia. Em 2018, o Seival FGS, por exemplo, rendeu 41% com modelos quantitativos seguidores de tendências, comemora o co-fundador Carlos Chaves. A Visia, com patrimônio sob gestão perto de R$ 900 milhões, busca ineficiências de mercado em repetição de padrões e usa aprendizado de máquina (machine learning) para executar ordens de compra sem impactar o mercado, explica o sócio Flavio Terni. Seu fundo Zarathustra rendeu 240% do Certificado de Depósito Interfinanceiro (CDI) em 2018.

A falta de familiaridade com conceitos e lógica das finanças colabora para a popularização dos robôs para aconselhamento. Para o CEO da Warren, Tito Gusmão, 85% dos investidores brasileiros preferem delegar decisões de investimentos por falta de conhecimentos suficientes sobre o tema.

A fintech, nascida em 2017, já atraiu mais de 50 mil clientes, maneja perto de R$ 300 milhões e usa a tecnologia para identificação de perfil e objetivos de investimento, recomendação e back office. Oferece quatro fundos próprios e, a partir de fevereiro, fundos de terceiros, graças à aquisição de uma corretora com recursos do aporte de um fundo do Vale do Silício. O serviço de gestão do portfólio de investimentos do cliente custa 0,5% ao ano, e produtos próprios têm taxa zero.

...

Já os serviços de gestão de patrimônio da Magnetis são sustentados por plataforma digital automatizada para entender o cliente, recomendar investimentos, definir carteiras e, se necessário, modificá-las. "A automatização cria eficiência operacional e deixa o serviço mais acessível", justifica o CEO Luciano Tavares. São 0,4% ao ano pela gestão e taxa administrativa de 0,3% a 0,5% para alocação nos fundos, com atendimento humano para os investidores.

A fintech Vérios, por sua vez, cobra entre 0,4% e 0,65% ao ano sobre o valor investido - somando custos como corretagem, custódia e outros, o valor total fica em 0,95% anuais. Robôs ajudam no gerenciamento da conta individual e no cálculo da alocação, com definição final pelo time da casa. No ano passado, o rendimento ficou entre 105% e 135% do CDI, detalha o CEO Felipe Sotto-Maior.”

Fora esse artigo nota-se por pesquisas na web que é muito grande a quantidade de projetos e artigos que tratam do aprendizado de máquina aplicado ao previsão de séries temporais de investimentos (ações, derivativos, criptomoedas).

Assim, fica comprovado que a utilização de sistemas de recomendação para investimentos é uma tendência forte no mercado financeiro, e um fundo de investimento gerenciado por algoritmos seria capaz de atrair muitos investidores.

A matriz CSD  permite  explorar  três  questões  essenciais  sobre  seu  projeto:  i)  as certezas, ou seja, o que já se sabe sobre o problema/desafio; ii) as suposições, ou seja, as hipóteses relacionadas ao problema/desafio; e iii) as dúvidas, de forma que ainda existem perguntas a serem respondidas.

Nessa matriz também é possível analisar o problema sob diferentes óticas, sendo as mais  comuns:  i)  atores,  que  são  as  pessoas  envolvidas  no  problema;  ii) cenários, onde ocorrem os problemas; e iii) regras, que representam a relação entre os atores e os cenários.

Tabela 1 Matriz CSD

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Certezas** | **Suposições** | **Dúvidas** |
| **Investidores e gestores de investimentos** | * Alta demanda por maior retorno para investimento | * Demanda por inovações no setor de fundo de investimentos | * Aceitação em delegar investimento para gestão por máquinas |
| **Aplicação de IA ao mercado financeiro** | * Avanços tecnológicos mudaram o mercado financeiro | * Capacidade dos algoritmos de prever os movimentos do mercado | * Geração de resultados consistentes de investimentos efetuados por robôs |
| **Regras** | * Utilização da tecnologia para apoiar o trabalho dos gestores de fundos | * Uso crescente da Inteligência Artificial na gestão de capital | * Fundos de investimento baseados em tecnologia vão dominar o setor financeiro |

1. Identificação da(s) pessoa(s) envolvida(s) no desafio:

A pessoa envolvida no projeto é o gestor de investimento, que todos os dias deve decidir o que fazer com seus investimentos: comprar, manter ou vender. Ele utiliza seus conhecimentos e ferramentas para avaliar e tomar as decisões, o que contribui para o seu sucesso ou não.

Caso ele possa contar com um sistema de previsão de preços que maximize suas chances de tomar executar decisões corretas e maximizar o valor do seus investimentos, ele poderá conseguir gerar retornos alfa, permitindo remunerar bem sua base de clientes e expandi-la por atrair novos investidores pela sua performance diferenciada.

1. Construção da proposta de solução:
   1. Requisitos da construção do protótipo/MVP, com descrição do experimento e as métricas de validação;
   2. Indicadores econômico-financeiros do projeto.

a. Requisitos da construção do protótipo/MVP, com descrição do experimento e as métricas de validação;

A construção de um protótipo/MVP permite tornar a oportunidade mais viável e tangível, de forma que constitui um passo crítico na criação do negócio. Por meio dele é possível ter um “embrião” do serviço, testar seu funcionamento e cumprimento de seus objetivos.

Os requisitos para o desenvolvimento do MVP seriam: obter dados do mercado de renda variável, algoritmos de aprendizado de máquina e infraestrutura computacional.

Os dados do mercado (dataset) incluem a variação dos preços nos pregões diários e informações econômico-financeiras que podem ser obtidas de diversas fontes: bovespa, yahoo, cvm, google. Tais fontes permitem fazer o download de arquivos csv. As informações serão utilizadas para o treinamento dos algoritmos de predição. O fornecimento de dados via API pela fonte é um atributo desejável, pois facilita o fluxo de informações, tornando a aplicação mais flexível e responsiva.

Os algoritmos são os modelos computacionais utilizados para a aprendizagem de máquina fornecidos por diversas bibliotecas abertas de machine learning, como scikitlearn, keras, tensorflow que compõem o ecossistema de Data Science do Python. Tais modelos devem ser implementados pela leitura e análise do dataset para gerar as funções de aproximação que serão usados na previsão dos preços. O treinamento consiste nessa etapa de implementação em que se é necessária uma infraestrutura de computadores para executar os métodos dos algoritmos.

A infraestrutura refere-se ao poder computacional necessário para executar o processamento dos dados pelos algoritmos, de forma a gerar o aprendizado. Quanto maior o volume de dados e a complexidade do algoritmo, maior necessidade de poder computacional (memória e processamento). Se forem utilizadas análises em tempo real, o que aumenta a velocidade dos dados, também é necessário mais poder. A utilização da computação em nuvem permite trabalhar com um custo e poder computacional variável, sendo ideal para o ambiente de desenvolvimento em que o MVP se insere. Algumas plataformas de cloud computing como, a Heroku, possuem pacotes de serviços gratuitos para desenvolvimento e teste da aplicação, além de garantir a escalabilidade para quando a aplicação estiver em produção.

A construção do MVP começa pelo planejamento das ações necessárias para o desenvolvimento da aplicação. Depois disso, o plano de ação fornecerá a orientação devida para avançar na execução do projeto.

1. 1. Levantamento dos dados
2. 2. Seleção de modelos de aprendizado de máquina
3. 3. Treinamento dos modelos e teste de previsão
4. 4. Classificação dos melhores modelos
5. Levantamento dos dados
   1. Yahoo: download CSV
   2. Google Finance + Pandas data-reader: via API
   3. Histórico de negociação dos últimos 5 anos ou mais
   4. Ações com alta liquidez que compõem o IBOVESPA
6. Seleção de modelos de aprendizado de máquina
   1. Pré-processamento dos dados
   2. Regression Decision Tree
   3. Random Forest
   4. Support Vector Machines
   5. Multi-layer Perceptron Regressor
   6. Long Short-Term Memory Neural Network
7. Treinamento dos modelos e teste de previsão
   1. Definição de parâmetros dos modelos
   2. Execução do treinamento
   3. Previsão e avaliação dos resultados
8. Classificação dos melhores modelos
   1. Ranking de acurácia dos modelos
   2. Seleção dos modelos mais eficazes

A validação do MVP deve ser feita por meio dos testes e análise do seu poder preditivo, quanto mais acurada a previsão de preço, mais o MVP atinge seu objetivo, que é provar ser capaz de prever o comportamento do mercado, e, portanto, ser capaz de recomendar investimentos.

Os testes podem ser efetuados em dois horizontes temporais. O primeiro seria o backtest que utiliza os dados históricos para fazer a previsão, e assim conferir se as previsões estariam corretas. Esse teste mostra o risco do underfitting, que significa que o modelo não foi capaz de prever corretamente os preços. O segundo seria testar com "dados futuros", ou seja, a medida que novos dados surgirem eles seriam usados para testar a previsão do modelo. Esse teste mostra o risco do overfitting, que significa que o modelo se especializou tanto nos dados de aprendizagem que não é capaz prever corretamente quando utiliza dados novos e diferentes, portanto não é útil.

Caso o MVP passe os testes simulados, uma outra forma de validação seria utilizá-lo em investimentos efetivos, o que demonstraria sua capacidade gerar resultados financeiros reais, e não somente simulados. Nesse aspecto algo que pode acontecer é que se os investimentos forem vultuosos, eles modem interferir na capacidade preditiva do modelo, pois poderiam gerar um novo padrão de comportamento que não havia sido incorporado ao modelo.

É importante ressaltar que o treinamento do modelo deve ser contínuo, mesmo que ele seja aprovado na validação, pois como as condições de mercado mudam e o modelo deve estar sempre se adaptando a elas, sob pena de perder seu poder preditivo.

Algumas métricas que podem ser usadas na validação do MVP são:

* Mean Absolute Error

A função mean\_absolute\_error calcula o erro absoluto médio, uma métrica de risco correspondente ao valor esperado da perda de erro absoluta ou perda de norma.

Se é o valor previsto da i-ésima amostra, e é o valor verdadeiro correspondente, então o erro absoluto médio (MAE) estimado é definido como

* Mean Squared Error:

A função mean\_squared\_error calcula o erro quadrático médio, uma métrica de risco correspondente ao valor esperado do erro ou perda quadrática (quadrática).

Se é o valor previsto da i-ésima amostra, e é o valor verdadeiro correspondente, então o erro quadrático médio estimado (MSE) é definido como

O MSE pode ser usado com Gradient Boosting regression para aprimorar a qualidade do modelo.

* R² score

R², o coeficiente de determinação fornece uma medida de quão bem as amostras futuras provavelmente serão previstas pelo modelo. A melhor pontuação possível é 1.0 e pode ser negativa (porque o modelo pode ser arbitrariamente pior).

Se i é o valor previsto da i-ésima amostra, e é o valor verdadeiro, então o R² é definido como:

O roadmap do negócio representa o planejamento da evolução do negócio ou aplicação. No caso do fundo de investimento, ele é determinado pelo aumento progressivo da complexidade e atividades executadas pelos algoritmos.

O primeiro passo é desenvolver o sistema de previsão de preços, e a progressão é rumo a realização das operações de compra e venda automatizadas criando um sistema completo para a gestão do fundo de investimento.

1. 1. Previsão de preços
2. 2. Alerta de oportunidades e riscos
3. 3. Monitoramento de investimentos realizados
4. 4. Geração de relatórios de performance
5. 5. Execução de ordens automatizadas

No caso do modelo preditivo, ele acompanharia a evolução do negócio incorporando mais dados, atividades ou algoritmos, de acordo com os testes e avaliações a serem feitos, os quais se pode antecipar sem a conclusão do MVP.

a. Indicadores econômico-financeiros do projeto.

Os indicadores econômico-financeiros são essenciais para avaliar a viabilidade de um projeto, de forma que se possam prever e monitorar seus custos e receitas, e evitar um projeto cujo custo seria superior ao valor gerado, configurando um desperdício de recursos (tempo, capital e pessoas).

A métrica base de avaliação do projeto seria o retorno sobre investimento (ROI), por meio dela seria capaz aferir a qualidade dos resultados obtidos com aplicação de aprendizado de máquina. Esse indicador pode ser utilizado tanto para o projeto e investimentos realizados com auxílio do modelo preditivo. Quanto maior o ROI do projeto maior a atratividade para seus sócios, e quanto aos investimentos, melhor para os clientes.

Outra métrica a ser utilizada é valor presente líquido (VPL), pois a avaliação somente em relação ao ROI não é suficiente já que ela é pode gerar um ROI alto em baixa escala e um ROI baixo em alto escala, e assim não demonstrar o valor monetário gerado como um todo. Por isso seria necessário utilizar também o VPL para aferir o valor presente do projeto e quantificar o seu ganho ou prejuízo dentro do prazo considerado. Se o ROI diminuir com o aumento da escala do projeto, mas o VPL ainda for positivo, o projeto mantém-se viável.

Não há como estimar o ROI e VPL antes da conclusão do MVP, pois sem ele não há perspectiva de receita nem gastos operacionais, por outro lado o custo de elaboração do MVP é considerado desprezível para o fluxo de caixa do negócio.

**DESCRIÇÃO DO DESENVOLVIMENTO DO TRABALHO**

1. Modelagem arquitetural do Sistema de Recomendação:
   1. Definição do(s) algoritmos;
   2. Definição da base de dados;
   3. Especificação dos resultados esperados e testes iniciais.

A construção do MVP requer que seja feita a modelagem arquitetural do protótipo, o que implica a definição do modelo, dos algoritmos de aprendizado e a base de dados.

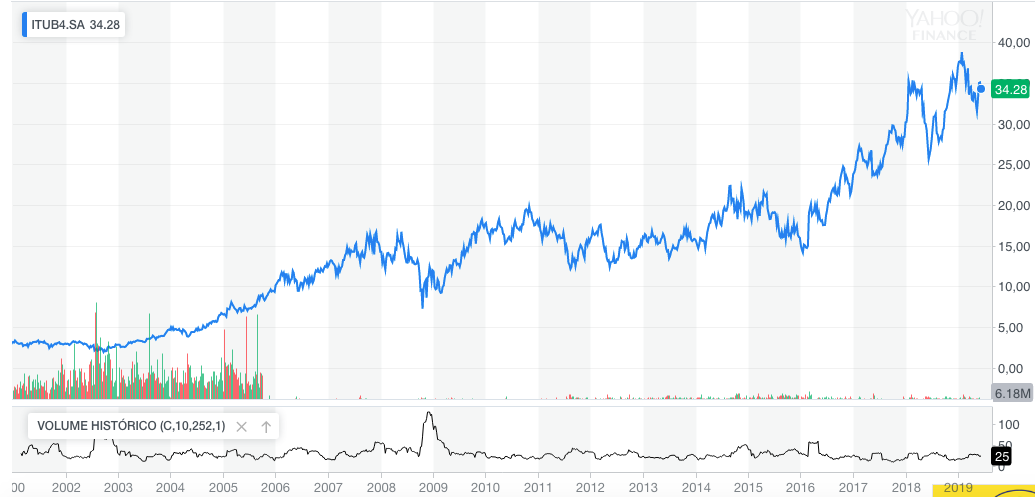
O modelo utilizado será a utilização das variáveis de preço (abertura, fechamento, máximo e mínimo) e volume para determinar se é recomendável vender ou comprar a ação. A determinação de compra ou venda é definida pelo acréscimo ou decréscimo do preço de fechamento em relação ao dia seguinte.

Os algoritmos pré-selecionados são: KNN, regressão linear e árvore de decisão regressiva.

A base de dados utilizado consiste nos preços históricos das ações PETR4 (Petrobrás), ITUB4 (Itaú) e EMBR3 (Embraer) . A escolha dessas ações deve-se pelo extenso histórico de preços com dados diários de negociação a partir do ano 2000. Além disso, o comportamento dos três ativos é diferente, isso permite testar se os algoritmos são capazes de detectar os diferentes padrões utilizados para treinamento.

Gráfico PETR4

Fonte: Yahoo finance

Gráfico ITUB4

Fonte: Yahoo Finance

Gráfico EMBR3

Fonte: Yahoo Finance

Os resultados esperados são que a arquitetura definida possibilite o treinamento dos algoritmos com na base de dados escolhida, de forma a obter um acerto razoável das previsões de compra ou venda.

1. Definição da principal abordagem técnica a ser utilizada no projeto.
   1. Seleção do(s) algoritmo(s) que será(ão) utilizado(s) na solução.
   2. Realização de testes iniciais com a base de dados.

A abordagem técnica utilizada é do aprendizado supervisionado usando uma classe binária (compra ou venda) como alvo da estimativa. Serão utilizados os dados até 31/12/2018 para treinamento e de 01/01/2019 até 31/05/2019 para teste.

Os testes iniciais utilizaram os algoritmos KNN, regressão linear e árvore de decisão regressiva, segundo o diagrama a seguir.

|  |
| --- |
| Diagrama de treinamentos |
|  |

Fonte: Orange3

Após o treinamento dos modelos foram feitas as avaliações, que apontaram os seguintes resultados.

|  |
| --- |
| Avaliação EMBR3 |
|  |

|  |
| --- |
| Avaliação ITUB4 |
|  |

|  |
| --- |
| Avaliação PETR4 |
|  |

O algoritmo árvore de decisão obteve o menor erro e maior R2, portanto foi o algoritmo com melhor capacidade preditiva dos três. Por conta disso, ele será o algoritmo usado para a solução proposta.

1. Implementação da solução utilizando algum framework de algoritmos de Machine Learning.
   1. Configuração do ambiente de desenvolvimento.
   2. Implementação ou seleção do(s) algoritmo(s).
   3. Validação e teste do algoritmo com dados reais.

Foi utilizado para realização dos testes iniciais o programa Orange3 disponível no pacote de aplicativos do Anaconda. O software implementa uma interface de programação visual para criação de treinamentos utilizando o framework Scikit-Learn. Portanto, a implementação da solução seria feita usando o SkLearn.

1. Avaliação da solução proposta.
   1. Definição da medida de acurácia a ser aplicada sobre os resultados gerados pelo(s) algoritmo(s).
   2. Realização de experimentos e coleta dos resultados.
   3. Ajuste de parâmetros.

A acurácia da previsão será medida pelo R² e MAE, de forma que as duas métricas permitirão aferir a qualidade do modelo pelo R² e o seu nível de acerto pela medida de erro MAE.

Foi feita uma análise de sensibilidade da árvore de decisão para buscar um melhor ajuste de parâmetros. Nos testes iniciais utilizou-se como tamanho para mínimo para os subsets a quantidade 5, notou-se que a redução desse valor para 3 aumentou a acurácia do algoritmo, conforme a tabela a seguir.

| **Métrica** | **5** | **3** |
| --- | --- | --- |
| MAE - EMBR3 | 0,037 | 0,031 |
| R² - EMBR3 | 0,853 | 0,874 |
| MAE - ITUB4 | 0,099 | 0,047 |
| R² - ITUB4 | 0,802 | 0,812 |
| MAE - PETR4 | 0,112 | 0,039 |
| R² - PETR4 | 0,777 | 0,845 |

**DESCRIÇÃO DO TRABALHO FINAL**

1. Fase de Preparação:
   1. Apresentação das ferramentas;
   2. Apresentação das técnicas;
   3. Apresentação das integrações;
   4. Especificação de algoritmos e frameworks utilizados;
   5. Apresentação das fontes de dados.

…

1. Fase de Processamento.
   1. Apresentação das ferramentas;
   2. Apresentação das técnicas;
   3. Apresentação das integrações;
   4. Especificação de algoritmos e frameworks utilizados;
   5. Apresentação das fontes de dados.

…

1. Fase de Armazenamento.
   1. Apresentação das ferramentas;
   2. Apresentação das técnicas;
   3. Apresentação das integrações;
   4. Especificação de algoritmos e frameworks utilizados;
   5. Apresentação das fontes de dados.

…

1. Fase de Visualização dos Dados.
   1. Apresentação das ferramentas;
   2. Apresentação das técnicas;
   3. Apresentação das integrações;
   4. Especificação de algoritmos e frameworks utilizados;

…

**RESULTADOS**

1. Descrição e análise dos resultados alcançados:
   1. Resultados positivos encontrados (caso existam). Explique;
   2. Resultados negativos encontrados (caso existam). Explique.

**CONCLUSÃO**

1. Apresentação da conclusão:
   1. Principais contribuições que seu projeto gera aos envolvidos;
   2. Inovações, particularidades ou vantagens que o projeto/resultado possui em relação a similares;
   3. Limitações do projeto;
   4. Próximos passos necessários para que o projeto evolua/se desenvolva.