

Alex Soares Prestes

Análise de fundos de investimentos utilizando grafo de visibilidade

Trabalho de conclusão de curso para obtenção do título de Bacharel em Física apresentado ao Programa de Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo.

Orientador: Gonzalo Travieso

São Carlos

1 de dezembro de 2023

RESUMO

Fundos de investimento compõem uma grande parcela dos ativos negociados na bolsa de valores, muitos destes são responsáveis por grandes empreendimentos, podem ser tão diversificados quanto se pode necessitar, e mesmo os que se enquadram como semelhantes cada gestor administra o fundo de formas diferentes. Buscando uma forma de aprofundar o processo de análise séries temporais financeiras, este trabalho tenta trazer uma alternativa para qualificá-las utilizando grafo de visibilidade como método para obter métricas e então ser aplicado como características do algoritmo de floresta de decisão aleatória. Usamos séries reais geradas de base de dados da CVM para investigar fundo de investimento imobiliários, e conseguimos diferenciar séries de maior e menor relevância.

Palavras-chave: Séries Temporais, Redes Complexas, Análise e Reconhecimento de padrões, Mercado Financeiro, Fundos de Investimento Imobiliário

SUMÁRIO

1	Introdução	4
1.1	Fundos de Investimento Imobiliários	4
1.1.1	Informes Periódicos	4
1.1.2	Avaliação de um fundo de investimento imobiliário	4
1.2	Redes Complexas e Séries Temporais	5
1.2.1	Série Temporal	5
1.2.2	Grafo de Visibilidade	5
2	Metodologia	6
2.1	Primeira Etapa	6
2.1.1	Obtenção dos Dados	6
2.1.2	Seleção dos Dados	7
2.1.3	Pré-Processamento	9
2.2	Segunda Etapa	9
2.2.1	Grafo de Visibilidade	9
2.2.2	Métricas do grafo de visibilidade	10
2.2.3	Construção do Fluxo de Caixa e Cálculo da TIR	10
2.3	Terceira Etapa	11
2.3.1	Outliers	11
2.3.2	Floresta aleatória	12
2.3.3	Buscando variáveis	12
2.3.4	Aplicando a Floresta de Decisão Aleatória e o grafo de visibilidade	13
3	Resultados	14
3.1	Variáveis mais relevantes	14
4	Conclusão	19
	REFERÊNCIAS	20

1 INTRODUÇÃO

Ao avaliar um Fundo de investimento é necessário analisar uma variedade de aspectos, desde o portfólio escolhido pelo gestor, como imóveis, certificados de recebíveis, ações e outros fundos de investimentos, também o quanto tem de retorno em cada um dos ativos escolhidos, e em alguns casos passa a ser necessário até saber quem são as empresas envolvidas em cada ativo. De forma geral essa avaliação é tão complexa quanto se queira, e a automatização desse processo é algo natural, a começar pelos passos mais simples.

1.1 FUNDOS DE INVESTIMENTO IMOBILIÁRIOS

Fundos de investimento imobiliário (FII) são ativos negociados na bolsa de valores que foram introduzidos em 1993 após a Lei 8668/93, na qual estabelece que o fundo trabalhe no regime de condomínio fechado, isto é, ao comprar uma cota só pode vendê-la para outro que queira comprá-la. A captação de recursos é feita por meio de emissões de novas cotas, tais quais são submetidas a novos compradores, junto com a proposta de como o gestor pretende utilizar esses recursos. Mesmo com essa proposta de uso dos recursos, muitas das vezes (quase sempre) a captação não é suficiente, nesses casos o gestor tem que realocar o patrimônio ou emitir empréstimo para cobrir o que falta.

1.1.1 Informes Periódicos

Mês a mês temos uma fotografia de como foi a movimentação contábil no fundo em um documento de informe mensal, este documento traz quanto do patrimônio está alocado em cada categoria de ativo, desde imóveis para aluguel, imóveis para venda ou até outros fundos de investimento e ações. No informe mensal é possível obter a sumarização dessas alocações, e no informe trimestral o detalhamento das mesmas, como geolocalização dos imóveis, especificação de quais fundos e ações o gestor está alocando patrimônio.

1.1.2 Avaliação de um fundo de investimento imobiliário

A taxa interna de retorno (TIR) é comumente utilizada na avaliação do fluxo de caixa (FC) de um investimento, à partir dela é possível obter uma ideia da rentabilidade e assim precificar um ativo. Essa taxa tem um papel fundamental nesse estudo, pois ela será o norte na análise. Como forma simplificar o uso e o entendimento dela nesse trabalho, o fluxo de caixa é obtido pelo próprio informe

mensal, onde o aumento patrimonial do fundo é considerada saída do caixa e os dividendos a entrada, a última entrada no caixa também tem o valor que o patrimônio do fundo esta avaliado.

1.2 REDES COMPLEXAS E SÉRIES TEMPORAIS

Uma rede complexa é em essência um grafo onde os nós e as arestas representam a dinâmica do sistema, o estudo de sistemas à partir de uma rede complexa permite a extração de uma variedade de métricas inerente a sua estrutura: número de nós, número de arestas, grau médio, menor caminho médio, coeficiente de distribuição dos graus e mais uma infinidade com cada uma trazendo diferentes informações sobre a rede. É interessante notar que essas medidas têm suas diferenças, mas o mais importante é que elas representam uma medida característica da rede ou por assim dizer uma assinatura da rede.

1.2.1 Série Temporal

Séries temporais são conjuntos de dados compostos de informações coletadas em diferentes momentos ao longo do tempo. Esses dados podem ser coletados em intervalos regulares ou irregulares, e são geralmente usados para entender e analisar fenômenos que mudam ao longo do tempo, como a evolução de vendas de um produto, o comportamento do mercado financeiro, ou até mesmo mudanças climáticas.

O objetivo da análise de séries temporais é encontrar padrões e tendências nos dados ao longo do tempo, o que pode ser usado para fazer previsões, detectar anomalias e entender como o fenômeno estudado se comporta. Para isso, são utilizadas técnicas estatísticas e matemáticas, como análise de autocorrelação, decomposição e modelagem de tendências e sazonalidade.

1.2.2 Grafo de Visibilidade

O grafo de visibilidade é um método que possibilita transformar uma série temporal em um grafo e assim analisá-la pela concepção de uma rede complexa, podendo obter suas medidas características. [1]

2 METODOLOGIA

A metodologia desse trabalho consiste em seguir uma versão simplificada e direta do processo knowledge-discovery in databases (KDD) [2], descoberta de conhecimento em base de dados, esse processo têm seus passos bem definidos, de forma que é extremamente simples dividir as tarefas de cada etapa e principalmente manter a consistência da metodologia aplicada.

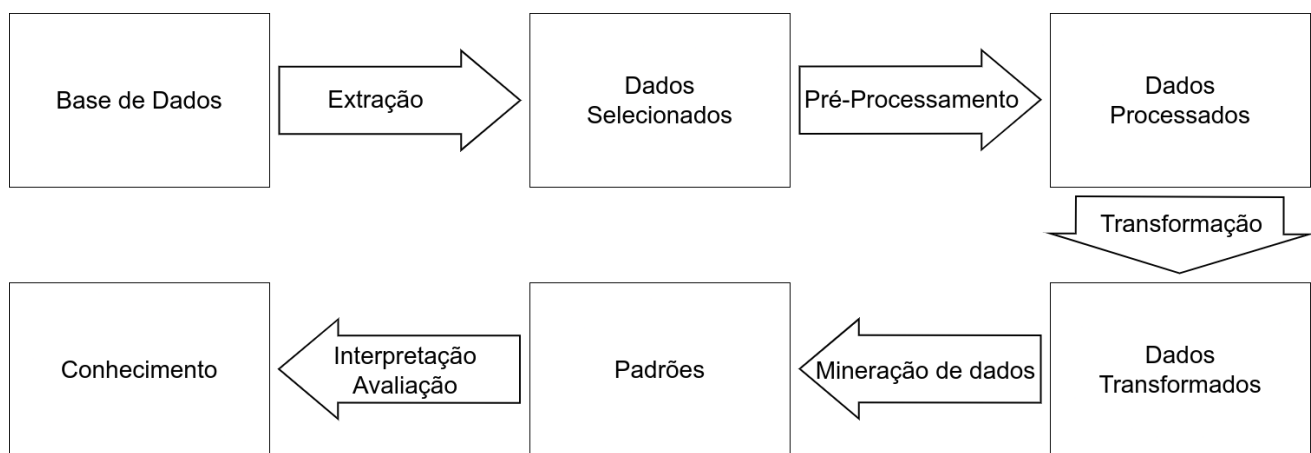


Figura 1 – Diagrama de cada etapa do processo KDD

De forma que possa descrever o processo da forma como foi executado, o processo KDD foi dividido em 4 etapas, as 3 primeiras referente aos métodos e a última parte representada pelo resultado e conclusão.

2.1 PRIMEIRA ETAPA

2.1.1 Obtenção dos Dados

Os dados utilizados nesse trabalho são obtidos diretamente do site da Comissão de Valores Mobiliários (CVM), onde os gestores são os responsáveis por fazer o envio dos relatórios. Os dados estão organizados em um arquivo de estrutura CSV compactados em um arquivo ZIP.

Portal Dados Abertos CVM

📁	../		
📄	inf_mensal_fii_2016.zip	25-Oct-2021 01:46	57K
📄	inf_mensal_fii_2017.zip	26-Dec-2021 10:21	407K
📄	inf_mensal_fii_2018.zip	25-Dec-2022 10:11	469K
📄	inf_mensal_fii_2019.zip	29-Jan-2023 10:51	551K
📄	inf_mensal_fii_2020.zip	29-Jan-2023 10:49	683K
📄	inf_mensal_fii_2021.zip	29-Jan-2023 10:47	844K
📄	inf_mensal_fii_2022.zip	29-Jan-2023 10:44	1004K

Figura 2 – Site de dados da CVM

A extração da base de dados foi feita por meio de um algoritmo em Python para automatizar o download, descompactação e leitura dos dados, assim podendo juntar todos os anos em uma única estrutura e obter as séries temporais para o período de 2016 até o início de 2023.

2.1.2 Seleção dos Dados

Conhecer os dados antes do pré-processamento é vital, saber que tipo de informação é esperada em cada campo e como é sua distribuição, são os primeiros passos para evitar erros na interpretação da análise. Apesar do foco ser a série temporal, nem todos os dados mudam no tempo, um exemplo é o cabeçalho do informe mensal, onde temos público alvo, a classificação, que determina onde o fundo tem que ter 2/3 do patrimônio investido e o prazo de duração, que quando é determinado é esperado uma maior distribuição de amortização.

Informe Mensal

Nome do Fundo:	TRX REAL ESTATE FUNDO DE INVESTIMENTO IMOBILIÁRIO - FII	CNPJ do Fundo:	28.548.288/0001-52 A
Data de Funcionamento:	15/10/2019	Público Alvo:	Investidores em Geral B
Código ISIN:	BRTRXFCTF003	Quantidade de cotas emitidas:	7.244.808,00
Fundo Exclusivo?	Não	Cotistas possuem vínculo familiar ou societário familiar?	Não
Classificação autorregulação:	Mandato: Renda Segmento de Atuação: Híbrido Tipo de Gestão: Ativa D	Prazo de Duração:	Indeterminado C
Data do Prazo de Duração:		Encerramento do exercício social:	Dezembro
Mercado de negociação das cotas:	Bolsa	Entidade administradora de mercado organizado:	BM&FBOVESPA
Nome do Administrador:	BRL TRUST DISTRIBUIDORA DE TÍTULOS E VALORES MOBILIÁRIOS S.A.	CNPJ do Administrador:	13.486.793/0001-42
Endereço:	Rua Iguatemi, 151, 19º andar- Itaim Bibi- São Paulo- SP- 01451-011	Telefones:	(11)31330350
Site:	www.brltrust.com.br	E-mail:	fii@brltrust.com.br
Competência:	12/2021		

Figura 3 – Cabeçalho Informe Mensal TRXF11 Referente ao período de 12/2021

Do cabeçalho foram selecionados apenas os dados que podem interferir na gestão do fundo.

O informe mensal traz também dados como o número de cotistas, mas como esse dado não é uma obrigação, nem todos os gestores fornecem e portanto será descartado.

Já os valores com respeito ao patrimônio, dividendos e despesas, esse são os dados que no decorrer desse trabalho serão transformados em séries temporais.

Data da Informação sobre detalhamento do número de cotistas¹	31/12/2021
Número de cotistas	39.634
Pessoa física	39.518
Pessoa jurídica não financeira	77
Banco comercial	
Corretora ou distribuidora	
Outras pessoas jurídicas financeiras	1
Investidores não residentes	7
Entidade aberta de previdência complementar	
Entidade fechada de previdência complementar	
Regime próprio de previdência dos servidores públicos	
Sociedade seguradora ou resseguradora	
Sociedade de capitalização e de arrendamento mercantil	
Fundos de investimento imobiliário	7
Outros fundos de investimento	23
Cotistas de distribuidores do fundo (distribuição por conta e ordem)	1
Outros tipos de cotistas não relacionados	

Figura 4 – Número de cotistas - Informe Mensal TRXF11 Referente ao período de 12/2021

2.1.3 Pré-Processamento

Antes de aplicar os métodos de classificação nos dados, é preciso fazermos um processo chamado de pré-processamento ou higienização, para tratar atributos com valores faltantes, outliers e as escalas diferentes para valores iguais, que podem afetar os resultados das análises. Os dados retirados do cabeçalho são do tipo categóricos, o processo de higienização desse tipo de dados é determinado pela moda, pois caso haja um erro em qualquer dos informes, predomina o valor com maior incidência, já nas séries temporais são tratados os valores nulos para zeros para evitar assim qualquer erro na fase de transformação.

2.2 SEGUNDA ETAPA

2.2.1 Grafo de Visibilidade

Com os dados selecionados e pré-processados, obter as séries temporais é apenas uma questão de aglomerar os valores de cada campo em uma conjunto ordenado pela data de referência, sendo por fim o objeto que será primariamente estudado. Cabe lembrar que também temos alguns dados categóricos que podem ser usados em uma classificação.

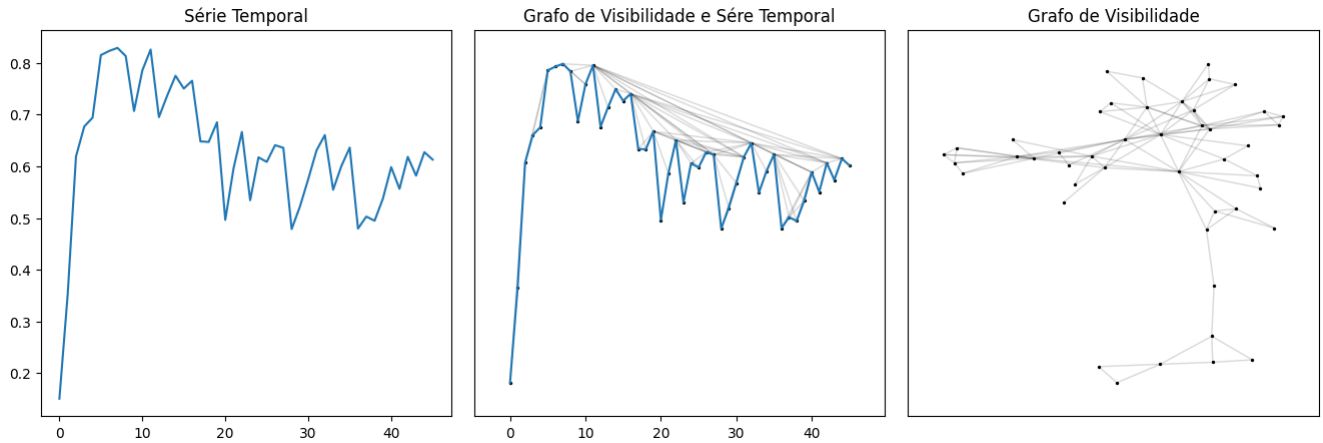


Figura 5 – Transformação para VG

A transformação da série temporal para grafo de visibilidade é feita tomando cada termo do conjunto como um nó e traçando uma reta até outros nós, caso não haja obstáculo, então a reta é definida como um vértice do grafo. [3]

$$y_c < y_b + (y_a - y_b) \frac{t_b - t_c}{t_b - t_a} \quad (1)$$

2.2.2 Métricas do grafo de visibilidade

A disponibilidade de métricas para o grafo de visibilidade são várias, abaixo uma lista com uma breve descrição das escolhidas. [4]

- **Número de nós:** quantidade de vértices no grafo;
- **Número de Arestas:** quantidade de ligações entre vértices;
- **Densidade:** a Densidade de uma grafo não-direcionado é dada por:

$$\rho = \frac{2m}{n(n-1)}$$

Onde m é o número de nós e n é o número de arestas;

- **Grau Médio:** média aritmética dos números de arestas em cada nó;
- **Coefficiente de Aglomeração:** é a razão da quantidade de triângulos pelo números nós, os triângulos são definidos com um conjunto de três nós conectados entre si;
- **Menor caminho médio:** média da menor distância entre todos os nós;
- **Coefficiente de distribuição dos graus:** Coeficiente angular da reta na distribuição dos graus.

Definidas as métricas podemos então gerar o vetor de características e assim padronizando a entrada de dados para o modelo que será treinado. [5]

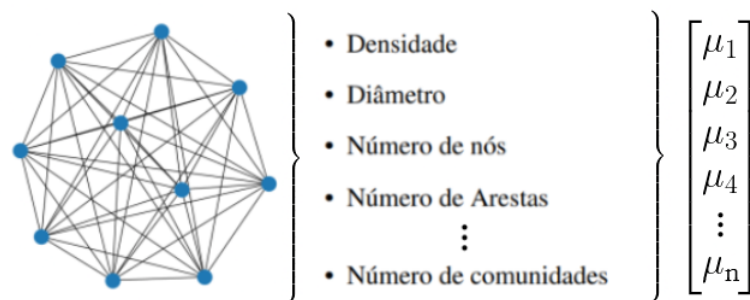


Figura 6 – Fonte: Gerada pelo autor

2.2.3 Construção do Fluxo de Caixa e Cálculo da TIR

O fluxo de caixa foi gerado a partir dos dados do informe mensal, cabe ressaltar que existe um sentido lógico no sinal dos valores, negativo indica a saída e positivo a entrada, como em essência queremos avaliar o fluxo de caixa do fundo, o primeiro valor será negativo indicando um investimento

do gestor no valor patrimonial do primeiro informe mensal, já os dividendos e amortizações serão entradas positivas, e por fim a última entrada será o valor patrimonial no último informe.

Data	Valor
01/01/2001	-R\$ 100,00
01/02/2001	R\$ 1,00
01/03/2001	R\$ 1,00
01/04/2001	R\$ 1,00
01/05/2001	R\$ 1,00
01/06/2001	R\$ 1,00
01/07/2001	R\$ 1,00
01/08/2001	R\$ 1,00
01/09/2001	R\$ 1,00
01/10/2001	R\$ 1,00
01/11/2001	R\$ 1,00
01/12/2001	R\$ 1,00
01/01/2002	R\$ 1,00
01/01/2002	R\$ 100,00

$$0 = \sum_{i=0}^n \frac{FC_i}{(1 + TIR)^{(DT_i - DT_0)/365}} \quad (2)$$

O cálculo da TIR é uma solução numérica da série (2), e assim a TIR será uma taxa de juros aproximada para o fluxo de caixa. O valor para a tabela acima é de 12.69% ao ano.

2.3 TERCEIRA ETAPA

2.3.1 Outliers

Para lidar com os valor discrepantes é necessário cuidados [6], isso porque eles não são inteiramente ruins, outliers podem ser parte crucial de uma base de dados, pois cada fundo tem algum valor discrepante em alguma variável e isso é o diferencial. Por outro lado, métricas como o coeficiente de distribuição de graus e a própria TIR precisam ser tratados, pois o cálculos para essas métricas são soluções numéricas e podem aparecer valores que não condizem com o esperado, para esses foi utilizado o método Z-Score para substituir os valores que fossem maior que 1.

Os outliers identificados no coeficiente de distribuição de graus e do cálculo da TIR foram substituídos por zero, a razão para essa troca é que o zero no coeficiente de distribuição é um valor neutro e fora da média dos valores padrão, e na TIR o zero significa que o investimento não teve retorno e nem prejuízo.

2.3.2 Floresta aleatória

Floresta aleatória ou Floresta de Decisão Aleatória é um algoritmo de aprendizado de máquina capaz de obter resultados sólidos com poucas modificações nos hiper-parâmetros. Esse algoritmo consiste em gerar n árvores de decisão e fundí-las para melhorar a predição e assim gerar árvores mais complexas, abrangendo uma maior diversidade de dados e estabilidade na previsão.

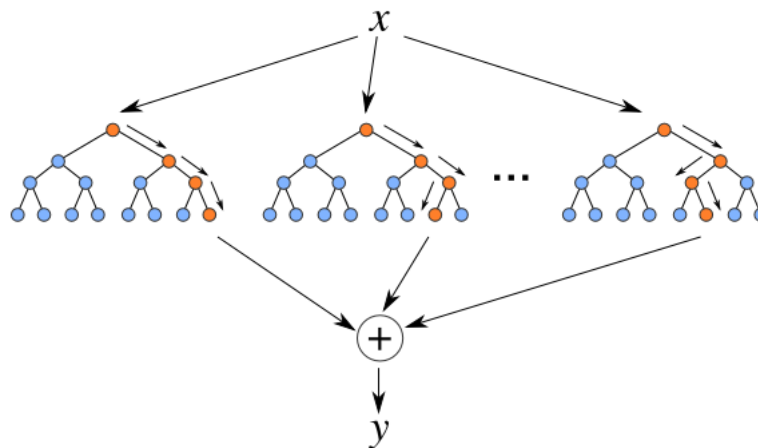


Figura 7 – Aspecto da Floresta de Decisão Aleatória - fonte: [7]

2.3.3 Buscando variáveis

Para obtenção das melhores variáveis, as que explicam melhor um FII no contexto da TIR, foi utilizado o algoritmo Floresta aleatória, esse modelo é usado para classificação e regressão supervisionada e uma de suas características mais interessante é a de obter o grau de importância das variáveis para o modelo, podendo assim selecionar as melhores mesmo em um conjunto de dados bastante complexo. O grau de importância é bastante simples de se interpretar, finalizado o treinamento do algoritmo teremos um conjunto de árvores de decisão, onde cada nó da árvore representa a tomada de decisão perante uma única variável, então é feita a contagem de nós por variável e em seguida a normalização desse valor.

2.3.4 Aplicando a Floresta de Decisão Aleatória e o grafo de visibilidade

variable	Valor_Ativo						
metrica	average_clustering	average_degree	average_short_path	coefficient_distribution_degree	density	number_of_edges	number_of_nodes
CNPJ_Fundo							
00.332.266/0001-31	0.812477	7.298701	3.529050	-0.934422	0.096036	281.0	77.0
00.613.094/0001-74	0.795282	12.025316	2.857514	-0.618819	0.154171	475.0	79.0
00.762.723/0001-28	0.805489	6.564103	4.695971	-1.040426	0.085248	256.0	78.0
00.868.235/0001-08	0.778299	7.948718	2.324009	-0.893427	0.103230	310.0	78.0
01.201.140/0001-90	0.801213	6.487179	3.329670	-0.701880	0.084249	253.0	78.0
...
49.184.940/0001-77	0.000000	1.333333	1.333333	0.000000	0.666667	2.0	3.0
49.318.661/0001-59	0.000000	1.000000	1.000000	0.000000	1.000000	1.0	2.0
49.489.928/0001-70	0.000000	1.000000	1.000000	0.000000	1.000000	1.0	2.0
97.521.194/0001-02	0.819705	7.540541	2.578304	-0.800385	0.103295	279.0	74.0
97.521.225/0001-25	0.759519	10.666667	2.861261	-0.674115	0.144144	400.0	75.0

Figura 8 – Corte da base de dados usada neste trabalho

Por último, teremos 6 características (métricas) para cada variável, como na Figure 8, essas serão as entradas para treinamento do algoritmo floresta de decisão aleatória, após o treinamento o algoritmo permite coletar as características e suas relevâncias, e retorna uma tabela com o nome da variável, métrica e valor da relevância, por fim agrupada-se por variável somando os valores das métricas como na Figura 9.

variable	metric	
Valor_Ativo	average_clustering	0.003913
	coefficient_distribution_degree	0.026237
	density	0.002200
	average_degree	0.001884
	average_short_path	0.010411
Total_Passivo	average_clustering	0.003183
	coefficient_distribution_degree	0.003261
	density	0.001548
	average_degree	0.072080
	average_short_path	0.007263

variable	
Total_Passivo	0.087336
Valor_Ativo	0.044644

Figura 9 – Obtenção dos resultados

3 RESULTADOS

Para obter o gráfico de erro da regressão foram treinados 10 modelos para cada menor tamanho de série da base de dados, isso é, o conjunto de dados é formado pela menor série até o tamanho máximo, isso porque quanto maior a série temporal, mais consistente serão as medidas do grafo de visibilidade. O que buscamos nesse primeiro gráfico é para qual comprimento a média do erro está mais próxima da mediana, esse será o comprimento utilizado para analisar os resultados.

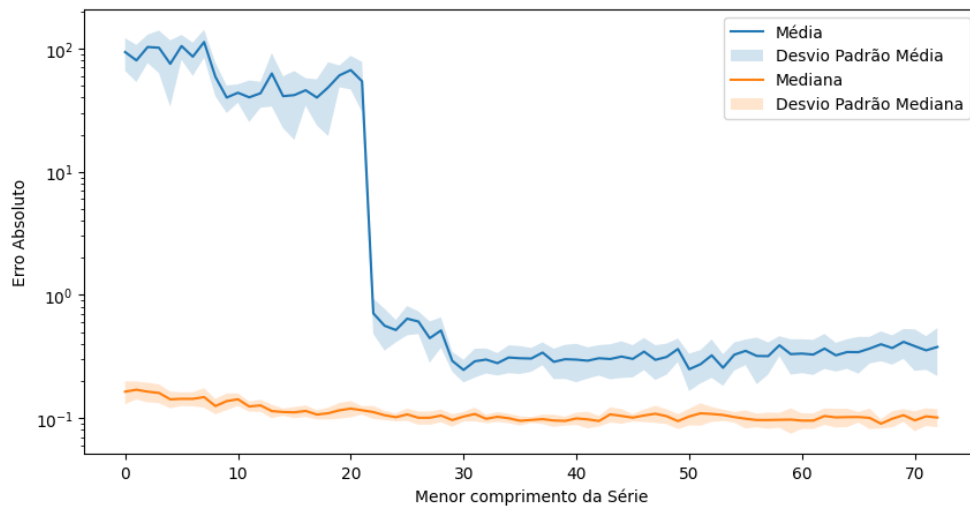


Figura 10 – Erro da base de teste do modelo Floresta aleatória

Avaliando a Figura 10, podemos ver que o melhor ponto para análise é com séries de comprimento maior que 30 elementos, com o intuito de evitar algumas divergências foi usado as séries com comprimento maior que 36 elementos.

3.1 VARIÁVEIS MAIS RELEVANTES

Após a regressão avaliar a relevância das variáveis é questão de interpretar os valores, e cabe lembrar que o contexto dessa regressão é quais características explicam o valor da TIR, basicamente estamos buscando quais variáveis são mais importante pra avaliar a rentabilidade do fundo e quais não tem significância para o explicar a TIR. O treinamento foi feito sem as variáveis utilizadas no fluxo de caixa, justamente para força o modelo a encontrar outras variáveis.

Para obter o gráfico abaixo, as variáveis foram agrupadas e os valores de cada métrica do grafo somadas, isso porque queremos avaliar a variável e não a métrica do grafo de visibilidade.

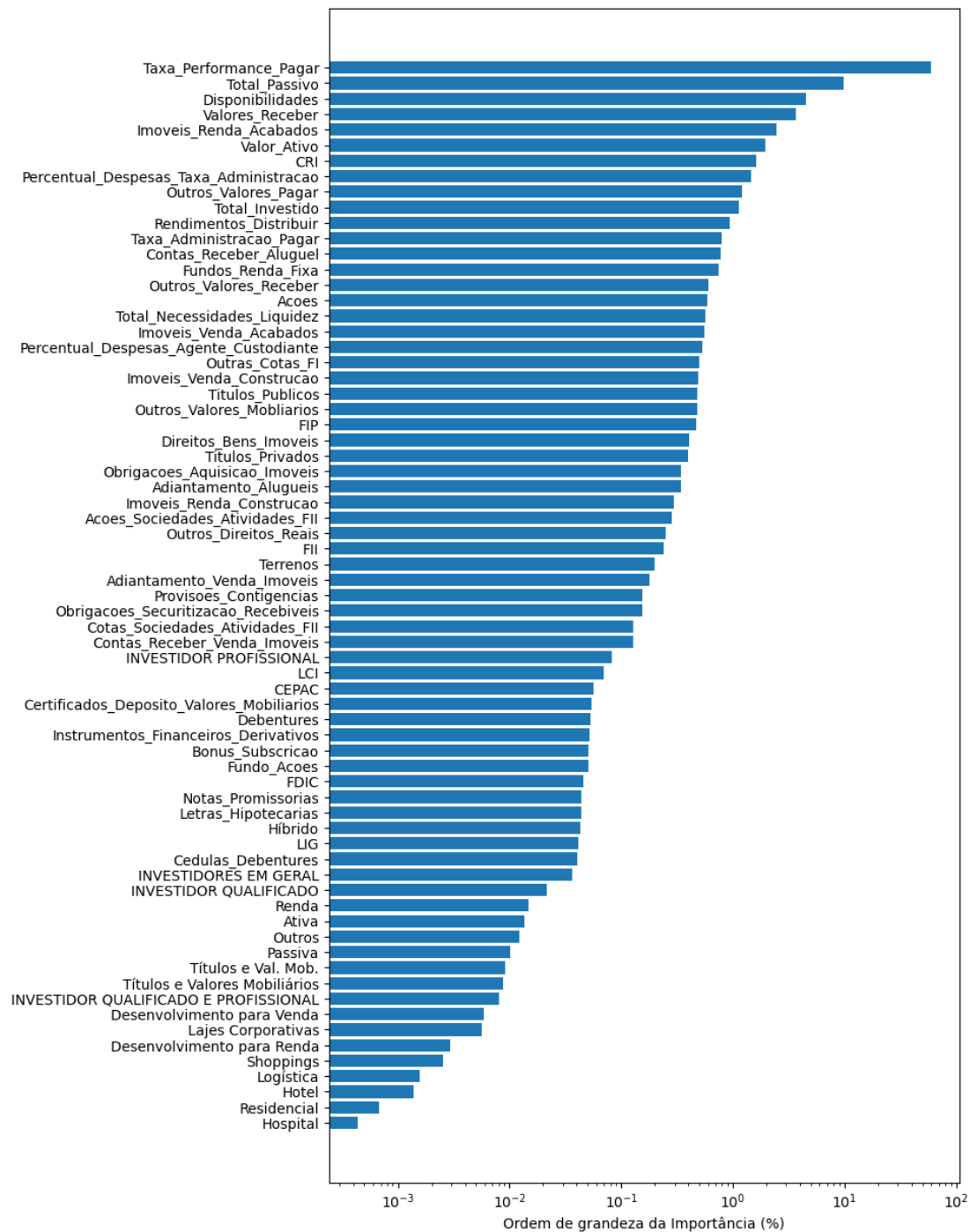


Figura 11 – Gráfico de importância das variáveis

Observar o gráfico acima por inteiro, não tem sentido, para isso vai ser feito duas análises, a primeira com as variáveis que tem mais de 1% de relevância e a segunda e com as características que somadas representam 1%.

Fazendo o corte nas variáveis que tem mais de 1% de relevância, observamos que a primeira já é a Taxa de Performance, essa variável é a que mostra quantos por cento o gestor recebeu de bônus

por atingir dividendos acima da meta.

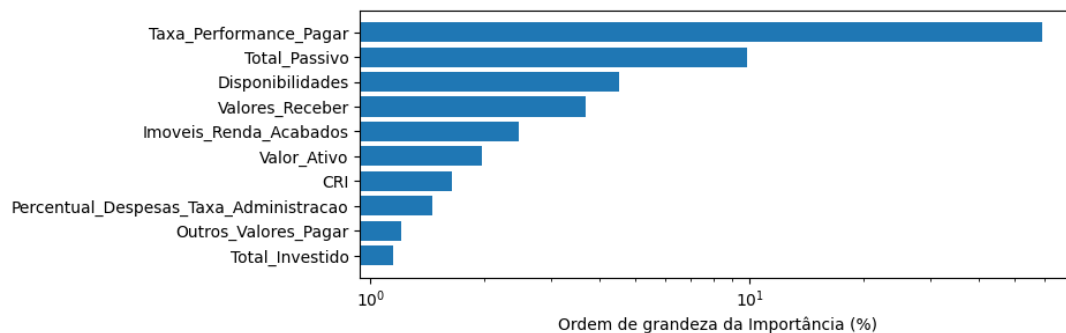


Figura 12 – Gráfico de importância das variáveis

- **Taxa de Performance a Pagar:** Valor que o gestor irá receber por atingir a meta de rentabilidade;
- **Total de Passivo:** Valor total a ser pago pelo fundo, varias variáveis estão sumarizadas nesse campo, desde despesas até os dividendos distribuídos;
- **Disponibilidade:** Valor que o gestor tem a disposição para pequenos gastos, costuma ser usado para pagamento de eventuais manutenção do imóvel, ou até para gastos com documentação;
- **Valores a Receber:** Valor total que o fundo irá receber, inclui aluguéis, venda, dividendos, multas entre outros;
- **Imóveis para Renda Acabados:** Essa característica é o valor que o fundo tem investido em imóveis para locação;
- **Valor de Ativos:** Valor total de todos os investimentos, todos os valores a receber e todo o valor em caixa;
- **CRI ou Certificado de Recebíveis Imobiliários:** Valor total investimento em CRI's, essa variáveis aparece justamente por representar uma parcela grande de FII que investem majoritariamente em CRI, costumam ter alta rentabilidade;
- **Despesas com a taxa de administração:** Valor gasto com a administração do fundo, no informe o campo é uma porcentagem em relação ao valor patrimonial;
- **Outros valores a pagar:** Totalizador dos gastos menores e eventuais;
- **Total Investido:** Valor total de todos os investimentos;

Como pode-se observar, todos os campos tem relação direta com gastos e retorno do fundo, mesmo as variáveis CRI e Imóveis para Renda Acabados, são características que os fundos têm maior valor investido.

Agora observando o segundo corte, das variáveis que somadas totalizam 1% de relevância, temos a informação mais importante dessa análise, isso porque quando se faz seleção de variáveis queremos tirar as que não tem interferem no modelo.

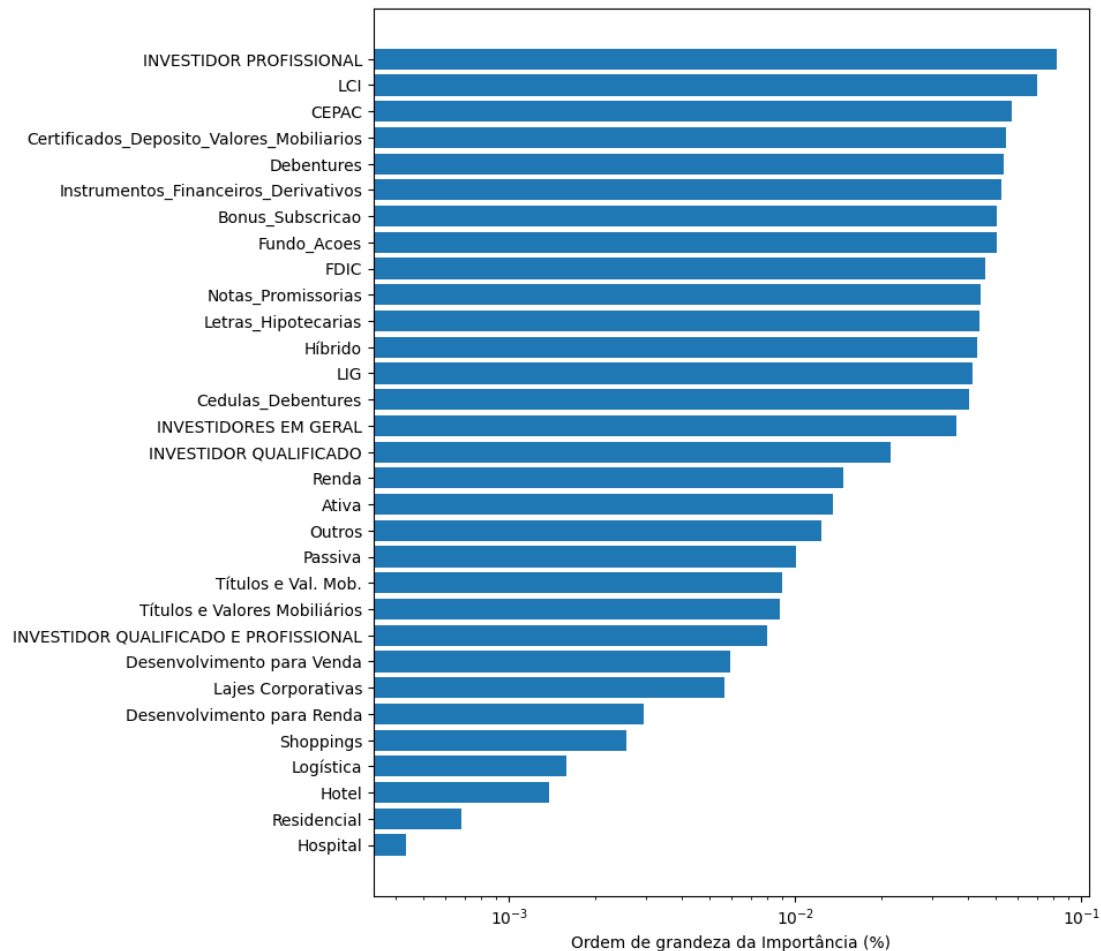


Figura 13 – Variáveis sem importância para o modelo

Aqui é necessário duas divisões, as primeira das variáveis categóricas e a segunda para descrever o porque das variáveis restantes

Era esperado a aparição dessas variáveis nessa posição, visto que a base de dados não é bem balanceada nesse sentido, fundos com segmento de atuação de hospital totalizam apenas 5, por exemplo, mesmo outras categóricas, tem uma representatividade muito baixa.

- **Público Alvo:** Investido Profissional, Investido Qualificado e Investidores em Geral;

- **Tipo de Gestão:** Ativa e Passiva;
- **Mandato:** Renda, Desenvolvimento para Venda, Desenvolvimento para Renda, Híbrido, Títulos e Valores Mobiliários;
- **Segmento de Atuação:** Shoppings, Híbrido, Lajes Corporativas, Outros, Logística, Hospital, Títulos e Val. Mob., Hotel, Residencial;

Partindo para explicar as variáveis seguintes, vou separar por classe de ativo, essa separação é para fazer mais sentido na explicação.

- **Ativos de Renda Variável:** Fundos de Ações, Instrumentos Financeiros Derivativos e FDIC;
- **Ativos de Renda Fixa:** Letras Hipotecárias, LIG, Notas Promissórias, Cédulas Debentures, CEPAC, Certificado Deposito Valores Mobiliários, Debentures;
- **Bônus de Subscrição:** Bônus recebido na forma de cotas, acontece quando um fundo ou ação distribui cotas para os acionistas;
- **LCI:** Título de renda fixa, utilizado por FII para caixa ou guardar sobras de emissões.

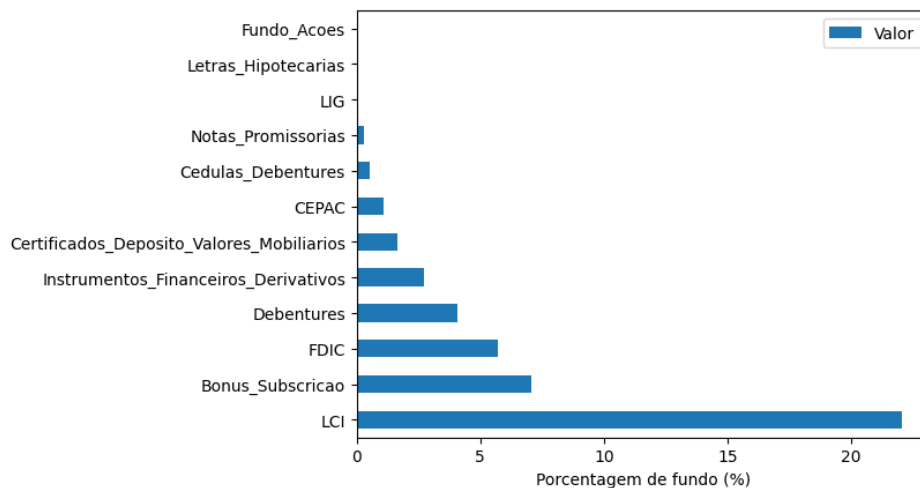


Figura 14 – Variáveis sem importância para o modelo

O gráfico acima é a porcentagem de FII que investem ou investiram nos ativos, podemos observar que é um valor extremamente baixo, mesmo a LCI que tem uma alta adesão, não é um ativo para quem quer rentabilidade, e sim segurança.

4 CONCLUSÃO

Fundos de investimento imobiliário tem uma estrutura bastante complexa, normalmente administrado por um gestor e sua equipe que são responsáveis por avaliar diversas opções de investimento para incorporar ao portfólio do fundo, isso torna cada FII único e de certa forma deixando a análise mais profunda.

O uso do grafo de visibilidade para decompor séries temporais financeiras em um vetor de medidas, faz com que possa extrair as características intrínsecas da série temporal, e por consequência as características de como cada gestor trabalha com o patrimônio do fundo.

Já a combinação do grafo de visibilidade com o algoritmo floresta de decisão aleatória, possibilita a extração das várias características, a mais importante é a facilidade de obter as variáveis de menos relevância para o modelo, e como mostra a análise são variáveis que tem baixa rentabilidade, ou não são ativos que costumam ser investidos pelos gestores, mostrando principalmente que é possível trazer um sentido lógico para os resultados.

Essa combinação de métodos mostra resultados significativos e possibilita análises de séries temporais por vários pontos de vista, como exemplo, se ao invés da regressão fosse feita a classificação do segmento de atuação dos fundos, seria então possível obter as variáveis que melhor descrevem o tipo do fundo de investimento.

E por fim cabe dizer que mais estudos no sentido de caracterização com dados reais são necessários, principalmente para compreender o domínio de atuação do grafo de visibilidade, isso porque é bem claro nos dados apresentado que as séries que tinham menos de 25 de comprimento causavam um perturbação muito grande no erro.

REFERÊNCIAS

- [1] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, “From time series to complex networks: The visibility graph,” *Proceedings of the National Academy of Sciences*, vol. 105, pp. 4972–4975, apr 2008.
- [2] G. Piatetski and W. Frawley, *Knowledge discovery in databases*. MIT press, 1991.
- [3] C. Bergillos, “ts2vg: Time series to visibility graphs.” <https://cbergillos.com/ts2vg/>, junho 2023.
- [4] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, “Characterization of complex networks: A survey of measurements,” *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.
- [5] F. A. Rodrigues and L. d. F. Costa, “Caracterização, classificação e análise de redes complexas,” Master’s thesis, Universidade de São Paulo, 2007.
- [6] “scikit-learn random forest regressor.” <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html#sklearn.ensemble.RandomForestRegressor>, junho 2023.
- [7] C. Baia, “Decision tree e random forest.” <http://carlosbaia.com/2016/12/24/decision-tree-e-random-forest/>, junho 2023.