

# Projeto de Pesquisa de Iniciação Científica

## Insper - Engenharia da Computação

**Aluno:** Vitor Grando Eller

**Professor:** Raul Ikeda Gomes da Silva

**Tema:** Aplicações de Redes Neurais Recorrentes e Fatores de Prémio de Risco no mercado de ações

**Título:**

### Introdução

Apesar da teoria proposta por Kendall defender que os movimentos dos preços do mercado são imprevisíveis [1], existe uma mobilização entre os estudiosos para construir modelos que sejam capazes de capturar possíveis tendências nos retornos. Tal tarefa é tratada com importância, uma vez que uma maior preditibilidade na precificação dos ativos representaria uma maior probabilidade de retorno, com redução nos riscos do investimento.

Um dos métodos concebidos foi o Modelo de Precificação de Ativos Financeiros (CAPM), o qual diz que o retorno em excesso é proporcional ao beta da ação. A partir deste, Fama e French propuseram em 1992 o Modelo de 3 Fatores [2], onde defendiam que, além do beta de mercado, o retorno em excesso seria explicado também por outros dois fatores relacionados à valores fundamentais das companhias.

Apesar disso, ambos modelos são lineares e não são eficientes na captação de todas as sinergias existentes entre as variáveis e saídas do sistema. Tal questão foi muito abordada no século XX, através da construção de modelos que utilizassem Redes Neurais.

As Redes Neurais, são sistemas computadorizados, que possuem como objetivo a simulação do processo do pensamento humano. Diferente dos modelos lineares, elas possuem a capacidade de examinar os dados após uma abstração de seus ruídos e distorções, sendo assim capaz de extrair relações não-lineares entre as variáveis preditivas.

Dessa forma, diversas pesquisas foram feitas para observar a usabilidade desta tecnologia no mundo das finanças. Wanjawa e Muchemi [3] propuseram um modelo baseado em Redes Neurais Artificiais (ANN) para prever o preço de ações nas bolsas, Pantazopoulos [4] utilizou Sistemas Neuro-Fuzzy para fazer estimativas de dados financeiros e criar estratégias de *trading*, e Batres-Estrada [5] avaliou a usabilidade de redes com Multicamadas de Perceptrons (MLP) em séries temporais financeiras de múltiplas variáveis.

Após isso, houve uma evolução no âmbito computacional que acarretou no desenvolvimento de técnicas capazes de analisar uma quantia maior de dados e com múltiplos níveis de abstração, sem a necessidade de supervisão humana. O *Deep Learning* trouxe inovações que possibilitaram a automatização da extração de representações complexas entre os dados, passando por uma estrutura hierarquizada de camadas, as quais recebem a ajuda de um algoritmo de retropropagação para compreender como a

máquina deve alterar seus parâmetros internos para computar as representações das próximas camadas.

A partir desta evolução, estudos surgiram a fim de comparar modelos lineares e não-lineares, a fim de confirmar as vantagens dos últimos, e avaliar qual seria a melhor técnica de *Deep Learning* quando da utilização no meio financeiro [6-9].

Porém, em todos os casos os estudos foram conduzidos em poucas ações e em amostras temporais com cenários econômicos constantes e bem definidos, além de utilizarem técnicas que, como exemplificado por Gamboa [12], não se encontram entre as mais propícias para se trabalhar com séries temporais.

Nos últimos anos houve o desenvolvimento de uma nova técnica de *Deep Learning*, as Redes Neurais Recorrentes (RNN). Tais redes incorporam a habilidade de analisar conexões ao longo de séries temporais, lançando mão de uma *Long Short Term Memory*. Essa habilidade torna as RNNs interessantes para o desafio de se prever o retorno de uma ação, dado que ações tomadas hoje no mercado podem impactar os ativos somente no futuro, sendo inclusive utilizada por Alberg e Lipton [11].

## Objetivos

O presente projeto apresentará uma análise de modelos de precificação de ativos que façam uso de fatores de prêmio, buscando analisar as diferenças entre modelos não-lineares e modelos clássicos, como o CAPM e o Modelo de 3 Fatores [2]. Para isso, utilizará duas classes de Redes Neurais, as Artificiais e as Recorrentes, visando também interpretar as desigualdades presentes nos resultados de tais classes. Além disso, buscará entender quais os custos relacionados ao uso de tais técnicas no dia-a-dia de um fundo de investimento.

## Metodologia

Fatores de prêmio de risco representam uma exposição à mercado para uma classe específica de ações. Fama e French[10] propuseram em sua pesquisa mensurar os fatores relacionados à diferença de performance de *Value Stocks* e *Growth Stocks*, obtendo dois novos fatores que explicavam o excesso de retorno de um ativo. Atingiram então uma equação linear relacionando a esperança de retorno com os prêmios de risco e os betas associados.

$$E[R_t^i] = E[R_t^m]\beta_m^i + E[R_t^{SMB}]\beta_{SMB}^i + E[R_t^{HML}]\beta_{HML}^i$$

Onde:

- $E[R_t^i]$  representa a esperança do retorno em excesso da ação  $i$  para o instante  $t$ ;
- $E[R_t^m]$  representa a esperança do retorno em excesso do mercado, para o instante  $t$ ;
- $\beta_m^i$  representa a exposição da ação  $i$  em relação ao mercado;
- $E[R_t^{SMB}]$  representa a esperança do retorno em excesso do fator *Small Minus Big* para o instante  $t$ ;
- $\beta_{SMB}^i$  representa a exposição da ação  $i$  em relação ao fator *Small Minus Big*;

- $E[R_t^{HML}]$  representa a esperança do retorno em excesso do fator *High Minus Low* para o instante  $t$ ;
- $\beta_{HML}^i$  representa a exposição da ação  $i$  em relação ao fator *High Minus Low*.

Logo, dados relacionados às operações das empresas analisadas, presentes em divulgações de resultado, serão coletados a fim de recriar esses fatores e o Modelo de 3 Fatores. Para testar a real efetividade dos modelos serão selecionadas uma variada gama de ações, vindas de empresas que pertencem a diferentes setores da economia, como sugerido por Alber e Lipton [11].

Após a construção do modelo linear, os fatores serão padronizados, a fim de eliminar discrepâncias relacionadas aos tamanhos das companhias analisadas. Será possível então construir os modelos não-lineares, implementando a ANN e a RNN. As informações também serão separadas em amostras de treinamento e teste, com o intuito de não permitir o *overfitting*. Por se tratarem de séries temporais os dados não poderão ser embaralhados durante essa separação, e para possibilitar análises de diversos cenários econômicos será preciso capturar amostras temporais de diferentes períodos tanto no conjunto de treinamento quanto no de teste.

Em seguida, serão testadas tanto a capacidade preditiva dos modelos quanto a viabilidade de utilizá-los na composição de carteiras de investimento. A primeira será testada através do uso de métricas estatísticas que validem as discrepâncias entre os valores preditos de retorno e de volatilidade e os valores reais, e a segunda com a criação de um portfólio hipotético e a mensuração dos retornos, aliados à índices financeiros, deste. Além disso, a carga computacional do Deep Learning será medida, a fim de calcular qual o real custo de implementação desta técnica no mundo real.

Por fim, as análises descritas acima serão unidas, com o intuito de compreender se as vantagens trazidas pela computação acabam por sobrepor os gastos ocasionados pela mesma.

## Resultados Esperados

Utilizando o *backtesting* e a composição de carteiras hipotéticas, serão produzidos resultados comparativos entre os modelos, a fim de validar se modelos não-lineares realmente possuem um maior poder preditivo quando comparado aos modelos lineares. Além disso, também será validado se os modelos que utilizam *Deep Learning*, possuem alguma aplicabilidade operacional real na composição de carteiras de investimento.

## Referências

- [1] Kendall, M. G., Hill, A. B. (1953). The analysis of Economic Time-Series-Part I: Prices. *Journal of the Royal Society of Statistics, Series A*, 116, 11-34.
- [2] Yu, T. H. K., & Huarng, K. H. (2008). A bivariate fuzzy time series model to forecast the TAIEX. *Expert Systems with Applications*, 34, 2945-2952.

