|  |
| --- |
| **Universidade Católica de Moçambique**  **Faculdade de Economia e Gestão**  **Cadeira**: Inteligência Artificial  **Estudante**: Mauro de Assunção Peniel  **Medidas de desempenho de um algoritmo no aprendizado de máquina**  **Docente**: Eng Francisco Chimbinde  Beira  2023 |

**Resumo**

Os algoritmos de aprendizado de máquina são um ramo da inteligência artificial que permitem que os sistemas aprendam e melhorem com o tempo sem serem explicitamente programados. O trabalho acadêmico abordou o tema das medidas de desempenho de um algoritmo no aprendizado de máquina, com o objetivo de avaliar a eficácia de diferentes medidas na identificação do desempenho de um modelo de aprendizado de máquina.

Foram selecionadas quatro medidas comuns de desempenho: acurácia, precisão, recall e F1-score, que são amplamente utilizadas na avaliação de modelos de aprendizado de máquina. O estudo envolveu a aplicação de diferentes modelos de aprendizado de máquina em conjuntos de dados variados, com o objetivo de analisar como cada medida se comporta em diferentes situações.

Os resultados mostraram que a escolha da medida de desempenho mais adequada depende do tipo de problema e dos objetivos da aplicação. A acurácia é uma medida simples e geral de desempenho, que pode ser útil em problemas de classificação binária, enquanto a precisão e o recall são mais adequados para problemas em que os erros de classificação têm custos diferentes.

O F1-score é uma medida que leva em consideração tanto a precisão quanto o *recall*, oferecendo um equilíbrio entre as duas medidas. Foi constatado que essa medida pode ser especialmente útil em problemas de classificação desbalanceados, nos quais a distribuição das classes não é uniforme.

Por fim, o estudo ressaltou a importância de selecionar cuidadosamente as medidas de desempenho de um modelo de aprendizado de máquina, tendo em vista que a escolha errada pode levar a interpretações equivocadas sobre o desempenho do modelo.

**Palavras Chaves:** Algoritmos, Aprendizado de Máquina, Inteligência artificial

**Abstract**

Machine learning algorithms are a branch of artificial intelligence that allow systems to learn and improve over time without being explicitly programmed. The academic work addressed the subject of performance measures of an algorithm in machine learning, with the objective of evaluating the effectiveness of different measures in identifying the performance of a machine learning model.

Four common performance measures were selected: accuracy, precision, recall and F1-score, which are widely used in evaluating machine learning models. The study involved applying different machine learning models to different datasets, with the aim of analyzing how each measure behaves in different situations.

The results showed that the choice of the most appropriate performance measure depends on the type of problem and the objectives of the application. Accuracy is a simple, general measure of performance that can be useful in binary classification problems, while precision and recall are more suitable for problems where classification errors have different costs.

The F1-score is a measure that takes into account both precision and recall, offering a balance between the two measures. It was found that this measure can be especially useful in unbalanced classification problems, in which the distribution of classes is not uniform.

Finally, the study highlighted the importance of carefully selecting the performance measures of a machine learning model, given that the wrong choice can lead to misinterpretations of the model's performance.

**Keywords**: Algorithms, Machine Learning, Artificial Intelligence

**Lista de Abreviaturas**

* SVM - Support Vector Machines
* RNAs - Redes Neurais Artificias
* KNN – K-Nearst Neighbors
* ML – Machine Learning

**Lista de Figuras**

[Figura 1 - 3 -](#_Toc130161979)

[Figura 2 - 4 -](#_Toc130161980)

[Figura 3 - 4 -](#_Toc130161981)

[Figura 4 - 5 -](#_Toc130161982)

[Figura 5 - 6 -](#_Toc130161983)

[Figura 6 - 7 -](#_Toc130161984)

[Figura 7 - 8 -](#_Toc130161985)

[Figura 8 - 9 -](#_Toc130161986)

[Figura 9 - 10 -](#_Toc130161987)

[Figura 10 - 11 -](#_Toc130161988)

**Índice**

[**1.** **Introdução** - 1 -](#_Toc130161854)

[**2.** **Medidas de desempenho de um algoritmo** - 2 -](#_Toc130161855)

[**3.** **Algoritmos usados na etapa de Aprendizado de Maquina** - 3 -](#_Toc130161856)

[**3.1.** **Regressão Linear** - 3 -](#_Toc130161857)

[**3.2.** **Arvores da Decisão** - 3 -](#_Toc130161858)

[**3.3.** **K-Nearest Neighbors (KNN)** - 4 -](#_Toc130161859)

[**3.4.** **Redes Neurais Artificias** - 5 -](#_Toc130161860)

[**3.5.** **Support Vector Machines (VSM)** - 5 -](#_Toc130161861)

[**3.6.** **Naive Bayes** - 6 -](#_Toc130161862)

[**3.7.** **Clustering** - 7 -](#_Toc130161863)

[**3.8.** **Regras de Associacao** - 8 -](#_Toc130161864)

[**3.9.** **Random Forest** - 9 -](#_Toc130161865)

[- 10 -](#_Toc130161866)

[**3.10.** **Gradient Boosting** - 10 -](#_Toc130161867)

[**4.** **Tipo de Aprendizado de Máquina** - 12 -](#_Toc130161868)

[**5.** **Conclusão** - 13 -](#_Toc130161869)

[**6.** **Referencia. Bibliográfica** - 14 -](#_Toc130161870)

# **Introdução**

A área de *Machine Learning* tem sido cada vez mais explorada em diversos campos do conhecimento, desde a análise de dados até a automação de processos complexos. Nesse contexto, a avaliação do desempenho dos algoritmos é uma etapa fundamental para garantir a eficácia do modelo desenvolvido. Este trabalho acadêmico tem como objetivo apresentar uma análise das medidas de desempenho de um algoritmo de *Machine Learning* implementado em *Python*. Para tanto, serão discutidos conceitos relacionados a avaliação de modelos de aprendizado de máquina e apresentadas as principais métricas utilizadas para avaliar o desempenho de um modelo, como precisão, *recall, F1-Score*, entre outras. Além disso, serão utilizados exemplos práticos para demonstrar como aplicar essas métricas em um modelo de *Machine Learning* em *Python*. O objetivo final deste trabalho é contribuir para uma melhor compreensão das medidas de desempenho de algoritmos de *Machine Learning* e seu papel na construção de modelos mais eficientes e precisos.

# **Medidas de desempenho de um algoritmo**

As medidas de desempenho de um algoritmo são utilizadas para avaliar a eficiência e a eficácia de um determinado método em resolver um problema específico. Existem várias medidas que podem ser utilizadas para avaliar o desempenho de um algoritmo, dependendo do tipo de problema e da aplicação em questão.

Algumas das medidas de desempenho mais comuns incluem:

* **Acurácia (Accuracy)**: A acurácia é a medida mais simples e direta do desempenho de um modelo. Ela representa a proporção de previsões corretas em relação ao total de previsões feitas pelo modelo. A acurácia é adequada quando as classes têm distribuições semelhantes e não há custo significativo associado aos erros de classificação.
* **Precisão (Precision)**: A precisão mede a proporção de exemplos classificados como positivos que realmente são positivos. É uma medida útil quando o custo de um falso positivo é alto.
* **Recall (Recall)**: O *Recall,* também conhecido como taxa de verdadeiros positivos (TPR), mede a proporção de exemplos positivos que são corretamente identificados pelo modelo. É uma medida útil quando o custo de um falso negativo é alto.
* **F1-score**: O F1-score é a média harmônica da precisão e do *recall,* sendo uma medida útil quando se deseja um equilíbrio entre ambas as medidas.
* **Área sob a curva ROC (AUC-ROC):** A AUC-ROC é uma medida que mede a capacidade de um modelo em distinguir entre classes. Ela representa a probabilidade de que o modelo classifique um exemplo positivo aleatório com uma pontuação maior do que um exemplo negativo aleatório.
* **Log-loss**: A log-loss é uma medida que penaliza previsões incorretas com um maior custo. É comumente usada em problemas de classificação binária em que as classes não são equilibradas.
* **Matriz de Confusão**: A matriz de confusão é uma tabela que mostra o número de previsões corretas e incorretas do modelo para cada classe. Ela é útil para avaliar o desempenho do modelo em termos de falsos positivos e falsos negativos.

# **Algoritmos usados na etapa de Aprendizado de Máquina**

## **Regressão Linear**

A regressão linear é uma técnica estatística usada para modelar a relação entre uma variável dependente (também chamada de variável resposta) e uma ou mais variáveis independentes (também chamadas de variáveis explicativas ou preditores). É uma técnica comum de aprendizado de máquina que é usada para prever valores contínuos, como preços de imóveis, demanda por produtos, vendas, etc.

Existem dois tipos principais de regressão linear: a regressão linear simples e a regressão linear múltipla. A regressão linear simples é usada quando há apenas uma variável independente, enquanto a regressão linear múltipla é usada quando há várias variáveis independentes.

* Exemplo de um pseudocódigo em python:

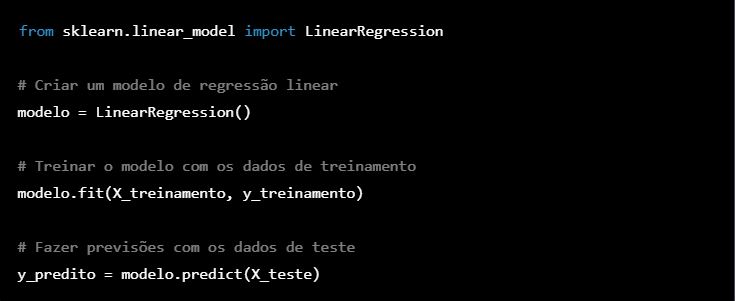


Figura 1

## **Arvores da Decisão**

A árvore de decisão é um algoritmo de aprendizado de máquina usado para modelar relacionamentos entre variáveis ​​independentes e dependentes, a fim de fazer previsões. O algoritmo é baseado em uma estrutura hierárquica de decisões em forma de árvore, onde cada nó representa uma decisão e cada ramificação representa uma possível consequência dessa decisão.

Em resumo, o algoritmo de árvore de decisão tenta dividir os dados em subconjuntos cada vez menores, com base nas características das variáveis, até que as observações sejam classificadas em um rótulo ou categoria específica. Durante o processo de construção da árvore, o algoritmo usa medidas de impureza, como índice de Gini ou entropia, para determinar qual variável independente é a mais importante na separação dos dados.

* Exemplo de um pseudocódigo em python:

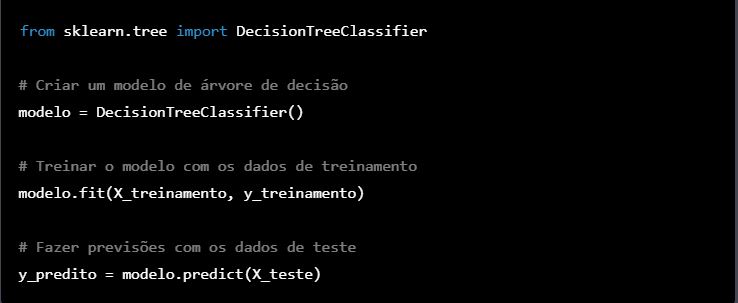


Figura 2

## **K-Nearest Neighbors (KNN)**

O algoritmo *K-Nearest Neighbors* (KNN) é um método de classificação e regressão usado em aprendizado de máquina. Ele é baseado no conceito de que instâncias semelhantes tendem a ter rótulos similares ou valores numéricos próximos. Em outras palavras, ele usa a similaridade entre as instâncias para prever o rótulo ou o valor de uma nova instância.

* Exemplo de um pseudocódigo em python:

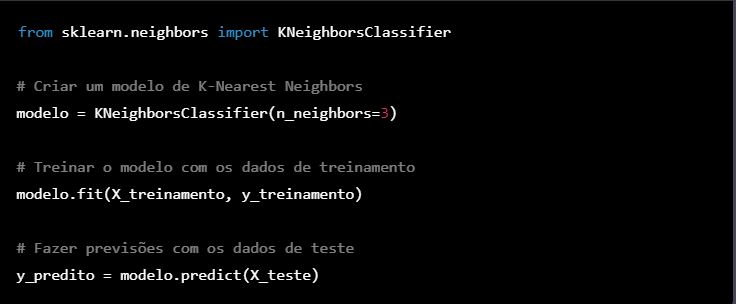


Figura 3

## **Redes Neurais Artificias**

Redes neurais artificiais (RNAs) são um tipo de algoritmo de aprendizado de máquina inspirado na estrutura e função do cérebro humano. Elas são compostas por várias unidades de processamento interconectadas, chamadas de neurônios artificiais, que são organizados em camadas. Cada camada pode ter uma ou várias unidades de neurônios.

As RNAs são capazes de aprender padrões e relações complexas entre as variáveis de entrada e saída de um problema, a partir dos dados de treinamento. Durante o processo de treinamento, os pesos das conexões entre os neurônios são ajustados para minimizar uma função de erro que mede a diferença entre a saída prevista e a saída real.

As RNAs podem ser usadas para problemas de classificação, regressão e outras tarefas, como reconhecimento de voz e imagem, processamento de linguagem natural, entre outros. Elas são especialmente úteis para problemas em que a relação entre as variáveis é não-linear ou difícil de modelar com outras técnicas.

* Exemplo de um pseudocódigo em python:



Figura 4

## **Support Vector Machines (VSM)**

*Support Vector Machines* (SVM) é um algoritmo de aprendizado de máquina que é utilizado para tarefas de classificação e regressão. O SVM encontra um Hiper plano que separa os pontos de diferentes classes, ou que melhor se ajusta aos dados de regressão.

A ideia principal do SVM é encontrar o hiper plano que maximiza a margem entre as classes. A margem é a distância entre o hiperplano e os pontos mais próximos de cada classe, chamados de vetores de suporte. A escolha do Hiper plano ótimo é feita de tal forma que a margem seja máxima, ou seja, os pontos mais próximos das classes sejam separados pelo maior espaço possível.

* Exemplo de um pseudocódigo em python:



Figura 5

## **Naive Bayes**

O algoritmo *Naive Bayes* é um modelo probabilístico baseado no teorema de *Bayes*, que é usado para classificação de dados em diferentes categorias. Ele é considerado "ingênuo" (naive) porque assume a independência entre as características dos dados, ou seja, que a presença de uma característica específica em um dado não afeta a presença de outras características.

O algoritmo usa uma base de treinamento para estimar a probabilidade de cada categoria, dada a presença ou ausência de determinadas características. Em seguida, usa essas probabilidades para classificar novos dados em categorias desconhecidas.

O *Naive Bayes* é frequentemente utilizado em aplicações de processamento de linguagem natural, como classificação de texto, análise de sentimentos e categorização de e-mails como spam ou não-spam. Ele também pode ser usado em outras áreas, como diagnóstico médico e reconhecimento de voz. Uma das principais vantagens do *Naive Bayes* é a sua simplicidade e rapidez de treinamento e classificação, mesmo com grandes conjuntos de dados.

* Exemplo de um pseudocódigo em python:

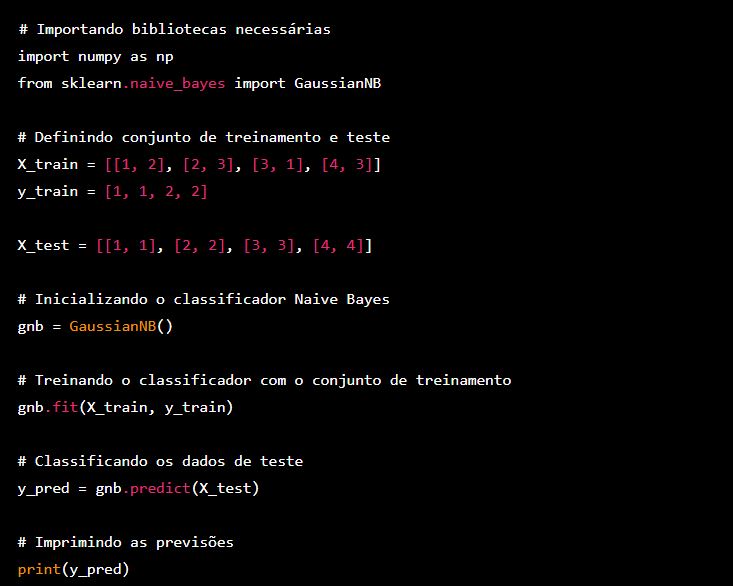


Figura 6

## **Clustering**

O algoritmo de *clustering* é uma técnica de aprendizado não supervisionado que tem como objetivo encontrar grupos ou clusters em um conjunto de dados. O objetivo é agrupar os dados em clusters de forma que os dados dentro de cada cluster sejam mais semelhantes entre si do que com os dados em outros clusters.

Os algoritmos de *clustering* são frequentemente usados em análise de dados, mineração de dados, reconhecimento de padrões, processamento de imagens e outras áreas. Eles podem ser usados para descobrir insights em grandes conjuntos de dados e identificar grupos de clientes, segmentos de mercado, entre outros.

* Exemplo de um pseudocódigo em python:



Figura 7

## **Regras de Associacao**

O algoritmo de regras de associação é uma técnica de mineração de dados que visa descobrir associações frequentes entre itens em um conjunto de dados. Ele é frequentemente usado em sistemas de recomendação e análise de mercado para identificar padrões de compra e preferências de consumidores.

O algoritmo de regras de associação mais conhecido é o Apriori. Ele usa um método de busca por força bruta para encontrar todas as regras de associação possíveis em um conjunto de dados. O algoritmo gera um conjunto inicial de itens frequentes e, em seguida, itera sobre o conjunto para encontrar regras de associação com base nas medidas de suporte e confiança.

O algoritmo de regras de associação é útil para descobrir relações ocultas e padrões em grandes conjuntos de dados e pode ser aplicado em uma ampla variedade de campos, incluindo marketing, comércio eletrônico, ciência de dados e pesquisa de mercado.

* Exemplo de um pseudocódigo em python:

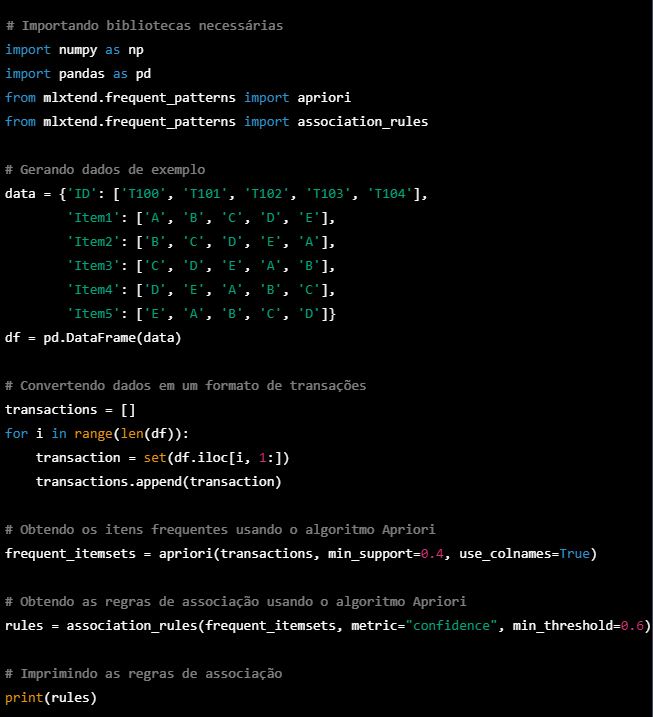


Figura 8

## **Random Forest**

*Random Forest* (Floresta Aleatória) é um algoritmo de aprendizado de máquina supervisionado que é amplamente utilizado para tarefas de classificação e regressão. Ele cria várias árvores de decisão aleatórias e combina as saídas dessas árvores para produzir uma predição final.

O *Random Forest* é um algoritmo poderoso e versátil, que é amplamente utilizado em uma variedade de tarefas de aprendizado de máquina, como classificação de imagens, detecção de fraude, previsão de mercado, entre outras.

* Exemplo de um pseudocódigo em python:

## 

Figura 9

## **Gradient Boosting**

*Gradient Boosting* é um algoritmo de aprendizado de máquina que é amplamente utilizado para tarefas de regressão e classificação. Ele constrói um modelo preditivo ao adicionar sucessivamente árvores de decisão simples ao modelo, de modo que cada nova árvore corrija os erros do modelo anterior.

O algoritmo *Gradient Boosting* começa criando uma única árvore de decisão para prever a variável alvo. Em seguida, ele calcula os resíduos entre as previsões da primeira árvore e os valores reais da variável alvo. Uma segunda árvore é então criada para prever esses resíduos, e assim por diante, até que um número desejado de árvores tenha sido criado.

* Exemplo de um pseudocódigo em python:

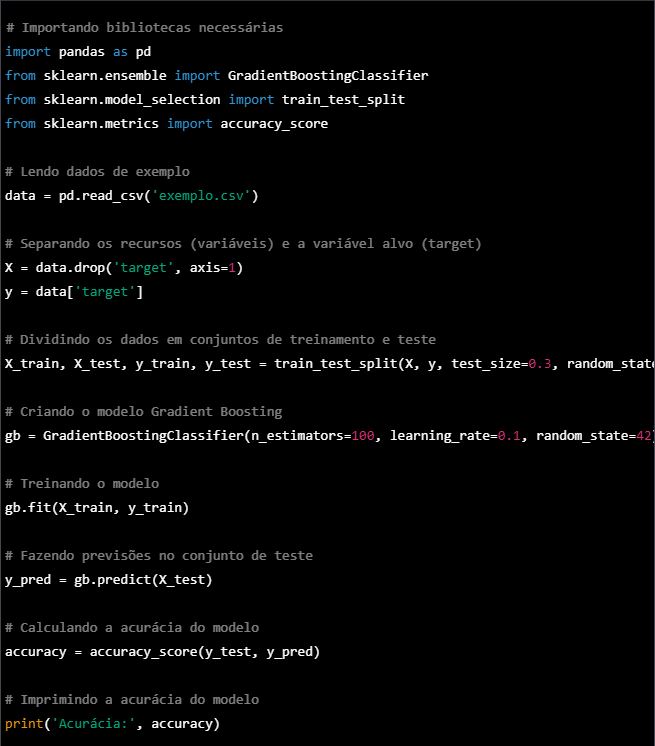


Figura 10

# **Tipo de Aprendizado de Máquina**

Existem três tipos principais de aprendizado de máquina:

* **Aprendizado Supervisionado**: Nesse tipo de aprendizado, o modelo é treinado com dados rotulados, onde as entradas estão associadas a uma resposta ou saída esperada. O objetivo é que o modelo aprenda a fazer previsões precisas para novas entradas. Exemplos de algoritmos de aprendizado supervisionado incluem regressão linear, árvores de decisão, redes neurais e algoritmos de classificação.
* **Aprendizado Não Supervisionado**: Nesse tipo de aprendizado, o modelo é treinado com dados não rotulados e o objetivo é encontrar estruturas e padrões nos dados. O modelo deve ser capaz de agrupar dados semelhantes e identificar padrões ocultos. Exemplos de algoritmos de aprendizado não supervisionado incluem *k-means*, análise de componentes principais e algoritmos de clusterização.
* **Aprendizado por Reforço**: Nesse tipo de aprendizado, o modelo é treinado para tomar decisões com base em um ambiente dinâmico e em mudança. O modelo aprende através da interação com o ambiente, recebendo feedback positivo ou negativo com base nas ações tomadas. O objetivo é maximizar as recompensas e minimizar as penalidades. Exemplos de algoritmos de aprendizado por reforço incluem *Q-learning*, SARSA e *Deep Reinforcement Learning.*

# **Conclusão**

Em conclusão, a análise das medidas de desempenho de um algoritmo é essencial para avaliar sua eficácia e eficiência em relação a outras soluções existentes. A seleção adequada de medidas de desempenho é crucial para garantir que os resultados obtidos sejam confiáveis e consistentes. Além disso, é importante considerar as limitações e suposições inerentes a cada medida de desempenho, bem como o contexto em que o algoritmo será aplicado.

Dentre as medidas de desempenho mais comuns, destacam-se a precisão, a *recall*, a  *F1-score*, a acurácia e a matriz de confusão. Cada uma dessas medidas oferece uma perspectiva diferente sobre a performance do algoritmo e é adequada para diferentes cenários.

No entanto, é importante lembrar que as medidas de desempenho devem ser usadas em conjunto com outras técnicas de avaliação, como validação cruzada e testes estatísticos, para garantir uma avaliação abrangente e rigorosa do algoritmo.

Por fim, é importante ressaltar que a escolha da medida de desempenho mais adequada depende do objetivo do estudo e do problema em questão. Portanto, é fundamental que o pesquisador tenha um conhecimento sólido sobre as diferentes medidas de desempenho e suas aplicações para realizar uma análise crítica e precisa do algoritmo em estudo.

# **Referencia. Bibliográfica**

* Alpaydin, E. (2010). *Introduction to machine learning.*
* Goodfellow, I. B. (2016). Deep learning. . *MIT press.*
* Hastie, T. T. (2009). The elements of statistical learning: data mining, inference, and prediction. *Springer Science & Business Media*.
* Shalev-Shwartz, S. &.-D. (2014). Understanding machine learning: From theory to algorithms. *Cambridge university press*.