1. **Objetivo**

Objetivo desse estudo é realizar a coleta de dados que envolvam os critérios de classificação dos jogadores da principal liga de basquete brasileira, a NBB (Novo Basquete Brasil), como foco específico na aplicação de técnicas de agrupamentos para identificar padrões e grupos no comportamento dos jogadores. Ao utilizar a análise de agrupamento, este trabalho tem como objetivo descobrir padrões que correlacionam os jogadores de basquetes analisando as seguintes características: por posição (armador, ala-armador, pivô, ala-pivô, ala); por critérios individuais, como: pontos, rebotes, assistências, arremessos, bolas recuperadas, tocos, erros, eficiência, duplos duplos. Os resultados contribuirão para uma melhor compreensão da estrutura tática que envolvem os times e dos fatores individuais que geram um entrosamento entre os jogadores.

1. **Coleta de dados**

Para este trabalho, realizei uma coleta de dados sobre a temporada 2022/2023 da NBB. O resultado dessa coleta de dados está disponível em:

<https://lnb.com.br/nbb/estatisticas/cestinhas/?aggr=avg&type=athletes&suffered_rule=0&season%5B%5D=71>

Esse banco de dados fornece uma coleção abrangente de estatísticas dos jogadores de basquete da liga nacional brasileira.

1. **Aprendizagem de máquina**

A aprendizagem de máquina, também conhecido como aprendizagem automática é uma área de pesquisa que tem como base o desenvolvimento de programas computacionais capazes de melhorar seu desempenho por experiência de forma automática, podendo ser compreendida segundo Marlon Vinícius Alves de Araújo como um método de análise de dados que se baseia em algoritmos que permitem extrair informações e identificar padrões nos dados, construindo automaticamente modelos analíticos para fazer previsões ou decisões. Tendo como ideia principal aprender a resolver problemas.

O Aprendizado de máquinas se dividi em alguns tipos, sendo os mais comuns:

* **Aprendizagem de máquina supervisionado:** é baseado em um conjunto de objetos para os quais as saídas desejadas são conhecidas, ou em algum outro tipo de informação que represente o comportamento que deve ser apresentado pelo sistema;
* **Aprendizagem de máquina não-supervisionado:** é baseado apenas nos objetos da base, cujos rótulos são desconhecidos. Basicamente, o algoritmo deve aprender a “categorizar” ou rotular os objetos.

1. **Aprendizagem de máquina não-supervisionado**

O aprendizado de máquinas não supervisionado é uma técnica de aprendizado que não necessita de um usuário supervisionando o modelo. Essa técnica é utilizada para explorar dados desconhecidos, quando não há categorização ou rotulação dos dados, aonde o algoritmo tenta entender por conta própria esse conjunto de dados e processa-los, agrupando esses dados de acordo com algumas semelhanças, encontrando estruturas e padrões ocultos em dados não rotulados. Essa forma de aprendizagem usa uma variedade de algoritmos para ajustar os dados em grupos amplos, agrupamentos e associações.

Aprendizagem não supervisionada segue algumas etapas básicas de tarefas, que são:

* Seleção de atributos
* Medida de proximidade
* Critério de agrupamento
* Algoritmo de agrupamento
* Verificação dos resultados
* Interpretação dos resultados

**3.1 Clustering**

Clustering ou agrupamento é uma técnica estatística usada para classificar elementos em grupos, de forma que elementos dentro de um mesmo cluster sejam muito parecidos, e os elementos em diferentes clusters sejam distintos entre si. O clustering, ou agrupamento, é uma técnica primitiva, pois nenhuma suposição é feita a respeito do número de grupos ou da estrutura dos grupos. O agrupamento é feito com base em semelhanças (similaridade) ou distâncias (dissimilaridades). As entradas necessárias são medidas de semelhança ou dados a partir dos quais as semelhanças podem ser calculadas. (VINÍCIUS, MARLON ALVES, 2021, p.17).

A ideia central da análise de cluster é a possibilidade de efetuar a classificação dos objetos em grupos, pois seu objetivo básico é descobrir agrupamentos naturais das variáveis. (VINÍCIUS, MARLON ALVES, 2021, p.17).

* 1. **Técnicas de cluster**

As técnicas de clustering (agrupamentos) tentam realizar em duas ou três dimensões tarefas que os observadores humanos não fazem tão bem, mesmo não tendo o conhecimento necessário de um agrupamento natural, são capazes de agrupar objetos em gráficos bidimensionais ou tridimensionais de forma visual.

Para realização desse trabalho, vamos consideram três métodos de agrupamentos distintos: K-means, AgglomerativeClustering e DBSCAN.

Análise de Agrupamentos

Tem como objetivo dividir os elementos da amostra em grupos, de tal forma que os elementos pertencentes a um determinado grupo específico sejam similares entre si, com base nas variáveis que foram medidas e os elementos em grupos distintos sejam dissimilares em relação as mesmas variáveis.