



08 a 11 de Outubro de 2018
Instituto Federal Fluminense
Búzios - RJ

USANDO O ALGORITMO K-MÉDIAS PARA CLASSIFICAR O PERFIL EMPREENDEDOR DOS ESTUDANTES DO INSTITUTO FEDERAL DO NORTE DE MINAS GERAIS (IFNMG) – CAMPUS MONTES CLAROS

Anne Oliveira Almeida¹ - annealmd@gmail.com

Débora Liliane Oliveira Borges¹ - deboraliliane81@gmail.com

Laércio Ives Santos¹ - laercio.ives@gmail.com

¹Instituto Federal do Norte de Minas Gerais - Montes Claros, MG, Brazil

Resumo. *This present study analyzes the entrepreneurial profile of the Computer Science and high school Technology students of the Instituto Federal do Norte de Minas Gerais (IFNMG) – campus Montes Claros. Given that the area of technology has as its main feature the innovation, leading to the emerging and rapid evolution of new markets and new ways of doing business, as well as the investment of the educational institution to develop an entrepreneurial profile in its students aiming that they leave the institution with competencies to act broadly in the labor market. The model used for the study, K-means, is an unsupervised machine-learning algorithm that groups individuals according to the study's focus characteristics - being a very efficient modeling method for this type of analysis.*

Keywords: *K-means, Entrepreneurship, Computer Science, clustering, Machine learning*

1. Introdução

Esse trabalho utiliza algoritmo K-means, também conhecido por K-médias e desenvolvido por Macqueen (1967), para mapear características Empreendedoras de alunos dos cursos de Ciência da computação e Técnico em informática do IFNMG-Campus Montes Claros.

Segundo Shapero (1975), o comportamento empreendedor tem como base: tomar iniciativa, organizar e reorganizar mecanismos sociais e econômicos a fim de transformar recursos e situações para proveito prático, aceitar o risco ou o fracasso.

Para Lima (2010) a personalidade empreendedora, começa a ser citada nos trabalhos sobre empreendedorismo somente a partir do final da década de 1990. Empreendedorismo é uma habilidade pessoal que motiva o indivíduo a colocar as suas ferramentas técnicas/pessoais e esforço na criação de novos produtos e serviços para corresponder uma demanda de mercado.

A área de tecnologia tem contado com apoio de empresas, iniciativa privada, na jornada de desenvolvimento de atributos empreendedores nos seus estudantes, visto que este perfil agrega valor a formação técnica e gera retorno acima da média para empregadores e sociedade.

Considerado como um fator chave para a sociedade contemporânea, o empreendedorismo estimula o indivíduo a identificar e entender as necessidades do seu meio e criar soluções que melhorem as condições de vida dos indivíduos, assim como atendam as suas exigências como potenciais clientes.

Segundo Bellotti et al (2012), além do valor pessoal, o empreendedorismo tem um potencial social porque a criação de um negócio exige a gestão de uma equipe de pessoas e a criação de postos de trabalho. O empreendedorismo pode contribuir para a coesão social, aumentando o emprego, o retorno financeiro e a satisfação no trabalho.

Atentos as necessidades do mercado e da sociedade, as Instituições de ensino vem se empenhando em formar profissionais com perfil empreendedor. Disciplinas que desenvolvem atributos empreendedores nos estudantes vem sendo incluídas no currículo, assim como atividades e competições que incentivam os estudantes ao autodesenvolvimento.

Dessa forma, estudar as características empreendedoras dos indivíduos e classifica-las é um meio de conhecermos pontos fracos e fortes para que se possa estudar alternativas para explorá-las melhor. Nesse sentido, utilizaremos o algoritmo K-médias que faz parte do espectro dos modelos não supervisionados do aprendizado de máquina para classificar alunos do Instituto Federal do Norte de Minas Gerais de acordo com características que são importantes no cenário do empreendedorismo.

O K-médias utiliza apenas dados de treinamento, a saída não é conhecida e sendo assim, não há dados para teste. Neste caso, o algoritmo deve por si só encontrar correlações entre os dados de entrada e agrupá-los homogeneamente.

Após a definição dos K centróides, o algoritmo agrupa os dados de acordo as suas similaridades e distâncias dos centros.

Segundo Dao et al (2015), o clustering (agrupamento) visa particionar a base de dados em clusters, de tal forma que os objetos no mesmo cluster são semelhantes e também diferentes dos objetos pertencentes a outros clusters. Esses requisitos são geralmente expressos por um critério de otimização e métodos de clustering consistem em encontrar o número (K) de partições mais adequado a base de dados foco do estudo, de forma que otimiza a classificação e separação dos objetos em clusters.

Para Aloise et al (2009), clustering é uma ferramenta poderosa para análise automatizada de dados. Alguns tipos de problemas que podem ser estudados: dado um conjunto de entidades, encontrar subconjuntos, ou clusters, que são homogêneos e/ou bem heterogêneos (separados). Existem vários critérios diferentes na literatura usados para expressar homogeneidade e/ou separação dos grupos a serem encontrados. Um critério bastante comum, que será usado neste estudo, é a soma mínima de distâncias euclidianas quadradas de cada entidade ao centróide do aglomerado o qual pertence, que expressa homogeneidade e separação.

No aprendizado de máquina, algoritmos de clustering, são muito utilizados em diversas áreas: como na genética, esporte, biblioteconomia, política, economia, estudos comportamentais, etc. Com o advento da internet e o avanço tecnológico, o seu uso para estudos mercadológicos, segmentação de clientes e colaboradores tem tido um importante papel no crescimento e lucro das empresas.

No esforço para segmentar os alunos do IFNMG levando em consideração as suas características empreendedoras, foi disponibilizado um questionário online para coleta de dados, estes foram posteriormente submetidos a métodos (Silhouette e WCSS- Within Cluster Sum of Squares) que identificam o K ideal para a base de dados em estudo, este K foi utilizado pelo algoritmo K-médias para formar os grupos de acordo com as variáveis escolhidas e posteriormente foi realizada a análise de cada agrupamento com a finalidade de traçar o perfil

empreendedor do grupo.

2. Metodologia

2.1 A metodologia utilizada pode ser subdividida nas seguintes etapas:

Coleta de dados dos estudantes através de aplicação de questionário, verificação da base de dados gerada, definição das variáveis a serem usadas no estudo e métricas de corte, aplicação de métodos específicos para análise de número de clusters adequado ao estudo e base de dados específicos, aplicação do método K-médias e finalmente análise dos clusters gerados pelo método K-médias.

2.2 Coleta de dados

Foi gerado um questionário on-line, este foi divulgado para os estudantes através de grupos fechados do IFNMG nas redes sociais e também pessoalmente na Instituição. Os estudantes foram informados, orientados e incentivados a participarem da pesquisa.

O questionário foi confeccionado de acordo a metodologia aplicada na disciplina de empreendedorismo e sob supervisão da docente responsável.

2.3 Verificação da base de dados

A ferramenta utilizada para coleta de dados gerou um arquivo .csv contendo uma tabela com todas as ocorrências e variáveis do questionário online.

2.4 Definição das variáveis a serem usadas no estudo e métricas de corte

De acordo com Bueno et al.(2004), o empreendedor é dotado de características, além dos atributos encontrados em administradores, os empreendedores são visionários, indivíduos que fazem a diferença, sabem explorar as oportunidades, são determinados e dinâmicos, dedicados ao trabalho, otimistas e apaixonados pelo que fazem, independentes e construtores do próprio destino, acreditam que o dinheiro é consequência do sucesso nos negócios, possuem liderança incomum, sabem construir uma rede de relacionamentos externos à empresa, planejam cada passo do negócio, possuem conhecimento, assumem riscos calculados e criam valor para a sociedade pela qual o empreendimento encontra-se inserido, em busca de soluções para melhorar a vida das pessoas.

Dessa forma, analisando as 16 variáveis do questionário, foi decidido o uso de 10 delas para a pesquisa, variáveis que teriam mais representatividade e impacto para o estudo de características empreendedoras dos alunos. As seguintes questões foram usadas para a classificação:

Aqui, o aluno se avaliou com uma nota de 1 a 5 para cada item:

- Proatividade na tomada de decisões;
- Persistência em resolver problemas;
- Toma riscos calculados (analisa tudo antes de agir);
- Hábil em resolver problemas e integrar soluções;

- Ciente de suas fraquezas e forças;
- Transmite integridade e confiabilidade;
- Sabe construir times e trabalhar em equipe;
- Disposição ao sacrifício para atingir metas.

Neste item do formulário o aluno poderia escolher uma opção dentre as seguintes:

1. Desde pequeno enxerga oportunidades para novos negócios;
2. Só abriria um negócio próprio se tivesse uma boa oportunidade;
3. Quer ser dono do próprio negócio para fazer as coisas exatamente do seu jeito;
4. Ter seu próprio negócio é a forma de encontrar o conforto e o padrão financeiro que deseja;
5. Só deseja ganhar dinheiro se seu trabalho melhorar a sociedade.

2.5 Análise do número de clusters

Foram utilizados dois métodos para definição do número de clusters ideal para a análise da base de dados.

Método Silhouette O principal desafio em tarefas de agrupamento é a análise e definição do número ótimo de grupos. O método Silhouette foi proposto por Rousseeuw (1987) para uso em agrupamentos por partição. A ideia é auxiliar a escolher o número ótimo de grupos e permitir que se construa uma representação gráfica do agrupamento encontrado.

Este método acusou a utilização de dois clusters ($K=2$) como melhor opção.

- $K = 2$: 0.4809972637442047
- $K = 3$: 0.4571394774620998
- $K = 4$: 0.4765129991160129

Método WCSS (Within Cluster Sum of Squares) - Método do Cotovelo Basicamente o que o método faz é testar a variância dos dados em relação ao número de clusters. Este testa até o momento que conforme o número de clusters aumenta não representa um valor significativo de ganho. Podemos ver esse “cotovelo” quando plotamos seus resultados em um gráfico. A partir do valor indicado pelo “cotovelo” no gráfico significa que não existe ganho em relação ao aumento de clusters. Nesse ponto, significa que o método encontrou o número ideal para o argumento K .

Para a base de dados utilizada para este estudo o WCSS definiu como melhor opção o uso de 4 clusters $K = 4$ para o agrupamento, como mostra o gráfico abaixo.

O gráfico 1 mostra que 4 clusters $K = 4$ é mais adequado para o agrupamento dos dados.

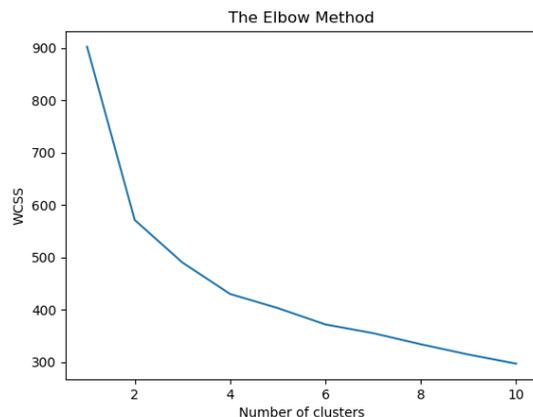


Figura 1- Elbow method (método do cotovelo)

3. Resultados

Aplicação do algoritmo K-médias para $K = 2$ e $K = 4$, gerando assim 2 gráficos de clusters. A figura 2 mostra a representação gráfica do algoritmo K-médias aplicado para 2 grupos. Mostrando os agrupamentos 0 e 1, assim como os seus centróides.

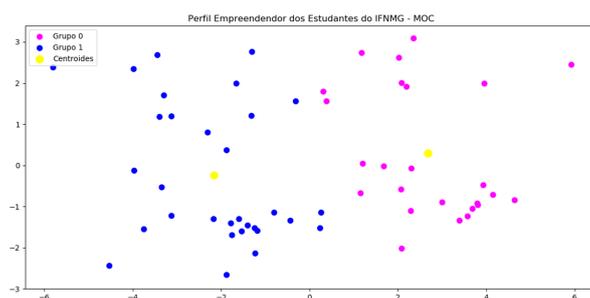


Figura 2- $K = 2$ (dois grupos/clusters)

A figura 3 mostra a representação gráfica do algoritmo K-médias aplicado para 2 grupos. A imagem mostra os agrupamentos 0, 1, 2 e 3 assim como os seus centróides.

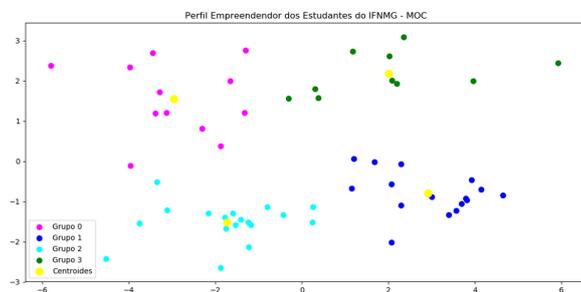


Figura 3- $K = 4$ (quatro grupos/clusters)

3.1 Análise dos clusters gerados pelo método K-médias.

K – 2 Para a análise das questões em que o aluno se avaliou com notas entre 1 e 5 foi definido que as variáveis avaliadas acima de 60% (notas 4 e 5) receberiam o rótulo de MUITO e abaixo de 60% (notas menores ou iguais a 3) de POUCO em relação a característica apresentada.

Tabela 1- Clusters K – 2

Cluster	Nº de Indivíduos	Porcentagem (%)
0	31	55,36%
1	25	44,64%

- **Grupo 0 (55,85 % da população total analisada):**

1. Proatividade na Tomada de Decisão – MUITO proativo (58,6% dos indivíduos desse grupo apresentam acima de 60% dessa característica).
2. Persistência em resolver Problemas – MUITO persistente (80,64% dos indivíduos desse grupo apresentam acima de 60% dessa característica).
3. Toma riscos Calculados (analisa tudo antes de agir) - MUITO (54,35% dos indivíduos desse grupo apresentam acima de 60% dessa característica).
4. Hábil em resolver problemas e integrar soluções – MUITO (70,96% dos indivíduos desse grupo apresentam acima de 60% dessa característica)
5. Hábil em se adaptar a novas situações – MUITO (74,19% dos indivíduos desse grupo apresentam acima de 60% dessa característica).
6. Ciente das suas Fraquezas e Forças – MUITO (80,63% dos indivíduos desse grupo apresentam acima de 60% dessa característica).
7. Transmite Integridade e Confiabilidade - MUITO (70,96% dos indivíduos desse grupo apresentam acima de 60% dessa característica).
8. Sabe construir time e trabalhar em equipe – MUITO (54,83% dos indivíduos desse grupo apresentam acima de 60% dessa característica).
9. Disposição ao sacrificio para atingir metas – MUITO (54,83% dos indivíduos desse grupo apresentam acima de 60% dessa característica).

Análise: Os indivíduos deste grupo se classificam como **MUITO EMPREENDEDORES**.

- **Grupo 1(44,64 % da população total analisada):**

1. Proatividade na Tomada de Decisão – POUCO (92% dos indivíduos desse grupo apresentam abaixo de 60% dessa característica).
2. Persistência em resolver Problemas – POUCO (80% dos indivíduos desse grupo apresentam abaixo de 60% dessa característica).
3. Toma riscos Calculados (analisa tudo antes de agir) - POUCO (92% dos indivíduos desse grupo apresentam abaixo de 60% dessa característica).
4. Hábil em resolver problemas e integrar soluções – POUCO (96% dos indivíduos desse

grupo apresentam abaixo de 60% dessa característica).

5. Hábil em se adaptar a novas situações – POUCO (88% dos indivíduos desse grupo apresentam abaixo de 60% dessa característica).

6. Ciente das suas Fraquezas e Forças – POUCO (88% dos indivíduos desse grupo apresentam abaixo de 60% dessa característica).

7. Transmite Integridade e Confiabilidade – POUCO (96% dos indivíduos desse grupo apresentam abaixo de 60% dessa característica).

8. Sabe construir time e trabalhar em equipe – POUCO (100% dos indivíduos desse grupo apresentam abaixo de 60% dessa característica).

9. Disposição ao sacrifício para atingir metas – POUCO (100% dos indivíduos desse grupo apresentam abaixo de 60% dessa característica).

Análise: Os indivíduos deste grupo se classificam como **POUCO EMPREENDEDORES**.

Tabela 2- Clusters *K* – 4

Cluster	Nº de Indivíduos	Porcentagem (%)
0	18	32,15%
1	18	32,15%
2	8	14,27%
3	12	21,47%

***K* – 4**

- Para a análise foi definido que a variável em questão seria avaliada segundo o critério percentual abaixo:
 - 0 a 40% : Muito pouco. (Notas 1 e 2)
 - 40% a 60% : Pouco (Nota 3)
 - 60 a 80% : Razoável. (Nota 4)
 - 80 a 100% : Desejável. (Nota 5)
- **Grupo 0 (32,15% da população total analisada):**
 1. Proatividade na Tomada de Decisão – muito pouco.
 2. Persistência em resolver Problemas – pouco.
 3. Toma riscos Calculados (analisa tudo antes de agir) - muito pouco.
 4. Hábil em resolver problemas e integrar soluções – muito pouco.
 5. Hábil em se adaptar a novas situações – muito pouco.

6. Ciente das suas Fraquezas e Forças – .muito pouco.
7. Transmite Integridade e Confiabilidade –muito pouco.
8. Sabe construir time e trabalhar em equipe – muito pouco.
9. Disposição ao sacrifício para atingir metas – muito pouco.

Análise: Os indivíduos deste grupo se classificam como **MUITO POUCO EMPREENDEDORES**, apresenta 90% das variáveis abaixo de 40% (Muito Pouco).

● **Grupo 1 (32,15% da população total analisada):**

1. Proatividade na Tomada de Decisão – Razoável.
2. Persistência em resolver Problemas – Razoável.
3. Toma riscos Calculados (analisa tudo antes de agir) - inconclusivo
4. Hábil em resolver problemas e integrar soluções – Pouco
5. Hábil em se adaptar a novas situações – Desejável
6. Ciente das suas Fraquezas e Forças – Desejável.
7. Transmite Integridade e Confiabilidade – Pouco.
8. Sabe construir time e trabalhar em equipe – Razoável.
9. Disposição ao sacrifício para atingir metas – Razoável.

Análise: Os indivíduos deste grupo se classificam como **RAZOAVELMENTE EMPREENDEDORES**, apresenta 30% das variáveis abaixo de 60%, 40% das variáveis entre 60% - 80% (Razoável) e 20% das variáveis acima de 80% .

● **Grupo 2 (14,27% da população total analisada):**

1. Proatividade na Tomada de Decisão – Pouco.
2. Persistência em resolver Problemas – Pouco.
3. Toma riscos Calculados (analisa tudo antes de agir) - Pouco.
4. Hábil em resolver problemas e integrar soluções – Muito pouco.
5. Hábil em se adaptar a novas situações – Pouco.
6. Ciente das suas Fraquezas e Forças – Pouco.
7. Transmite Integridade e Confiabilidade – Pouco.
8. Sabe construir time e trabalhar em equipe – Pouco.
9. Disposição ao sacrifício para atingir metas – Pouco.

Análise: Os indivíduos deste grupo se classificam como **POUCO EMPREENDEDORES**, apresenta 80% das variáveis abaixo de entre 40% e 60% (Pouco).

● **Grupo 3 (21,43% da população total analisada):**

1. Proatividade na Tomada de Decisão – Razoável.
2. Persistência em resolver Problemas – Desejável.
3. Toma riscos Calculados (analisa tudo antes de agir) - Desejável.
4. Hábil em resolver problemas e integrar soluções – Desejável.
5. Hábil em se adaptar a novas situações – Desejável.
6. Ciente das suas Fraquezas e Forças – Desejável.
7. Transmite Integridade e Confiabilidade – Desejável.
8. Sabe construir time e trabalhar em equipe – Desejável.
9. Disposição ao sacrifício para atingir metas – Desejável.

Análise: Os indivíduos deste grupo se classificam como **EMPREENDEDORES**, apresenta 80% das variáveis acima de 80% (Desejável).

3.2 Considerações

- **Empreendedor Situacional:** Grande parte dos estudantes expressou o seu desejo de empreender apenas se houvesse uma boa oportunidade de negócio.

Em ambos os grupos do $K - 2$, destaca-se a característica de empreender apenas caso haja uma oportunidade de negócio - grupo com características MUITO EMPREENDEDORAS (45,3%) e grupo com características POUCO EMPREENDEDORAS (40%). No caso do $K - 4$, essa característica está presente apenas no grupo com características MUITO POUCO EMPREENDEDORAS (55,55% dos indivíduos) e no grupo com características RAZOAVELMENTE EMPREENDEDORAS (77,7% dos indivíduos).

- **Empreender do “seu jeito” :** O estudante apontou como motivo para empreender a vontade de fazer as coisas do seu jeito, sem empregador.

Apenas o grupo com características MUITO EMPREENDEDORAS (20%) do $K - 2$ e o grupo com características MUITO POUCO EMPREENDEDORAS (27,78%) do $K - 4$ apresentam dados relevantes – acima de 20% - nesta variável.

- **Crescer em ambiente Empreendedor:** Este quesito investigou o número de estudantes que teve contato com o empreendedorismo no ambiente familiar e se sofreu ou sofre influência dessa experiência.

Apenas o grupo com características EMPREENDEDORAS (25%) do $K - 4$ apresenta relevância neste dado.

- **Empreender por conforto e retorno financeiro:** Foi apontando como motivo para empreender o retorno financeiro e o conforto proporcionado por este.

O grupo com características POUCO EMPREENDEDORAS (25%) do $K - 2$ e grupo com características POUCO EMPREENDEDORAS (62,5%) e grupo com características EMPREENDEDORAS (41,67%) do $K - 4$.

- **Empreendedor Social:** Alguns estudantes atrelaram o empreendedorismo ao bem estar social, apontando que um fator determinante para abrirem um negócio seria o impacto desse na sociedade.

Apenas o grupo com características POUCO EMPREENDEDORAS (37,5%) e o grupo com características EMPREENDEDORAS (25%) do $K - 4$ apresentam essa característica.

4. CONCLUSÃO

O algoritmo K-médias, de aprendizado de máquina não supervisionado foi eficiente na segmentação da base de dados utilizada.

Sendo o $K - 2$ apontado pelo método silhouete o que melhor expressou a heterogeneidade dos grupos e a homogeneidade interna dos seus elementos.

O $K - 4$, apontado pelo método wcss, foi mais eficiente na segmentação dos grupos, expressando características que não foram possíveis no anterior. Além de classificar os extremos, o $K - 4$ também segmentou grupos intermediários - favorecendo assim a uma análise mais complexa das características empreendedoras dos estudantes do IFNMG, campus Montes Claros.

Referências

- Aloise, Daniel and Deshpande, Amit and Hansen, Pierre and Popat, Preyas, "NP-hardness of Euclidean sum-of-squares clustering", *Machine Learning journal*, vol.75,2,245-248
- Bueno, Ana Maria and Leite, Magda L.G. and Pilatti, Luiz Alberto, "Empreendedorismo e comportamento empreendedor: como transformar gestores em profissionais empreendedores", XXIV Encontro Nac. de Eng. de Produção, 2004.
- Dao, Thi-Bich-Hanh and Duong and Khanh-Chuong and Vrain and Christel, "Constrained Minimum Sum of Squares Clustering by Constraint Programming", *Principles and Practice of Constraint Programming*, Pesant, Gilles (ed.), Springer International Publishing, Cham, 557–573
- Dornelas, José, "Empreendedorismo, Transformando ideias em negócios", LTC – Livros Técnicos e Científicos, 45-62;
- F. Bellotti and R. Berta and A. De Gloria and E. Lavagnino and F. Dagnino and M. Ott and M. Romero and M. Usart and I.S. Mayer, "Designing a Course for Stimulating Entrepreneurship in Higher Education through Serious Games", *4th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES'12). Procedia Computer Science journal*, 15, 174 - 186", 2012.
- J. Macqueen, "Some methods for classification and analysis of multivariate observations", In 5-th Berkeley Symposium on Mathematical Statistics and Probability, 281-297.
- Lima, Rosa Cristina Ribeiro AND Freitas, Ana Augusta Ferreira de, "Personalidade empreendedora, recursos pessoais, ambiente, atividades organizacionais e desempenho financeiro de empreendedores informais", *Revista de Administra, scielo*, vol.44, 04, 511 - 531.
- Nascimento, Pedro Paulo Marques do "Estudo de Técnicas de programação inteira no problema de monitoração não intrusiva de cargas residenciais", Universidade Federal do Rio de Janeiro, 2013, Tese de monografia.
- Rousseeuw, P., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.
- Shapero, Albert, "Entrepreneurship and economic development", 1975, p. 187.