**Projeto de Pesquisa e Planejamento de Atividades**

|  |  |  |
| --- | --- | --- |
| **Aluno**: Cleiton Pereira da Silva | | **Data início curso**: 17/03/2022 |
| **Orientador**: Leila Rabello de Oliveira | | **Defesa em:**  08/2023 |
| **Curso**: MBA em Data Science e Analytics - EaD 1º Sem /2022 | **Modalidade**:  Pós graduação | Turma:  DDSA 22 - 1 |

* **Título do projeto** *(Inicial)*

*Correlação entre Nota e Sensações Descritas em Avaliações de Filmes de Aventura em Franquias de Sucesso: Um Estudo de Data Analytics.*

*Sub-titulo:*

*Uma análise de Regressão Utilizando Variáveis Dummy para Investigar o Impacto das Sensações Descritas em Avaliações de Filmes de Aventura de Franquias de Sucesso na Nota do Filme.*

* **Introdução**

*(Contextualizar e apresentar a problemática do tema geral, ou seja, a importância do tema proposto e sua relevância. O texto deverá ser escrito de forma impessoal e toda informação utilizada deverá ser embasada por meio de trabalhos de fontes confiáveis com as devidas citações dos autores)*

*Quais sensações nos fazem avaliar positivamente uma aventura no cinema? Quais são as impressões cruciais para uma boa avaliação de um filme deste genero por um telespectador?*

*Dentro de uma avaliação de filme no site especializado IMDB (Internet Movie Database) é possível adicionar uma descrição breve com palavras-chave para contextualiza a nota atribuída para o filme avaliado e será investigando milhares das repostas deste banco de dados que traçaremos um paralelo entre as boas e más notas e as sensações descritas.*

*Antecipar estas impressões pode melhorar a diligencia dos produtores deste genero de filme para que alcassem a expectativa de seu publico sem contrariar o processo criativo artistico de suas produções, encaixando a ativação das sensações nos momentos mais oportunos através de tecnicas de filmagens e trilhas sonoras.*

*A base de dados a ser utilizada será do IMDB, que foi criado em 1990 como um banco de dados online de informações sobre filmes, programas de TV e celebridades. Atualmente, ele é uma das principais fontes de informação sobre cinema, com uma grande quantidade de avaliações e resenhas de usuários. As notas atribuídas pelos usuários aos filmes de aventura variam de 1 a 10, sendo que notas mais altas indicam uma avaliação mais positiva.*

*Com a ajuda de uma análise de regressão, será possível analisar milhares de indicações de uma amostra da base de dados do site IMDB e distinguir avaliações que contêm ou não determinadas sensações descritas por palavras-chave, como “angústia”, “alegria”, “liberdade”, e determinar se a presença ou ausência destas palavras-chave pode ser relacionada com avaliações com boas ou más notas.*

*A análise de regressão é uma técnica estatística utilizada para analisar a relação entre uma variável dependente (a nota da avaliação) e uma ou mais variáveis independentes (as palavras-chave que caracterizam determinadas sensações). Quando as variáveis independentes são categóricas, é necessário utilizar variáveis dummy, que representam cada categoria com uma variável binária (0 ou 1). Neste caso, indicar a presença ou ausência de determinada palavra-chave. Essas variáveis são importantes para explicar o efeito de cada categoria na variável dependente, como no caso das sensações descritas nas avaliações de filmes de aventura em franquias de sucesso afetando as notas atribuídas destas avaliações.*

* **Objetivo**

*(Qual o objetivo principal do trabalho, ou seja, qual pergunta deve ser respondida ao final da sua pesquisa)*

*Este estudo de data analytics identificará possíveis correlações entre as notas atribuídas às avaliações de filmes de aventura em franquias de sucesso e as sensações descritas nessas avaliações. Mais especificamente, busca-se antever estatisticamente quais sesações tendem a instigar notas positivas nas avaliações com base em sensações específicas que podem ou não estar logicamente ligadas às notas mais altas das avaliações, permitindo uma melhor compreensão de como as percepções dos espectadores estão relacionadas às notas atribuídas aos filmes.*

*Ao utilizar a técnica de análise de regressão com variáveis dummy em um banco de dados, que transformará palavras-chave descritas nas avaliações dos filmes em variáveis dummy, será possível identificar quais variáveis tendem a melhorar a nota da avaliação e quais tendem a piorar a nota da avaliação.*

*Essa abordagem pode ajudar a indústria cinematográfica a entender melhor como as opiniões e sensações dos espectadores influenciam a nota que um filme recebe, possibilitando ajustes na produção, marketing e distribuição de filmes de aventura em franquias de sucesso. Além disso, a análise pode fornecer informações valiosas para os tomadores de decisão, como produtores e distribuidores, que podem utilizar as informações para melhorar a qualidade dos filmes, das prévias dos filmes e aumentar a satisfação dos espectadores sem podar criativamente as atividades artisticas.*

* **Material e Métodos**

*(Descrever o(s) método(s) de coleta de dados e a(s) ferramenta(s) de análise a ser(em) utilizada(s) no trabalho de conclusão de curso, ou seja, como será a condução da pesquisa e a forma de obtenção dos resultados, por exemplo, fontes de dados, técnicas, procedimentos, índices, entre outros)*

Pesquisa bibliografica, definição das sensações, base de dados IMDB, filmes escolhidos para a analise.

*Para esta monografia em específico, o material utilizado será composto por avaliações de filmes de aventura em franquias de sucesso publicadas no site IMDB, que foram coletadas por meio do sistema Kaggle (Uma comunidade mundial de Machine Learning and Data Science).*

*A partir da transformação de palavras-chave das avaliações em um resumo da opinião central descrita, serão criadas variáveis dummy que representarão cada sensação identificada nas avaliações, as quais serão utilizadas para criar um modelo de regressão múltipla utilizando variáveis dummy. O modelo será criado com a finalidade de identificar, dentro da lista de filmes indicada, quais sensações descritas nas avaliações possuem correlação com notas mais altas, e quais possuem correlação com notas mais baixas.*

*A lista de filmes que comporão a base de dados é composta por os seis filmes da “terra média”, sendo três nomeados como “Senhor dos Aneis” e três como “O Hobbit”, e onze filmes do “mundo bruxo”, sendo oito nomeados como “Harry Potter” e três como “Animais fantásticos”.*

*Por fim, será realizada uma validação cruzada para avaliar a eficácia do modelo em prever as notas dadas às avaliações de filmes de aventura em outras franquias de sucesso com base nas sensações descritas nas avaliações. Essa validação cruzada será realizada utilizando uma amostra dos dados coletados que não foi utilizada para criar o modelo.*

* **Resultados Esperados**

*(Descrever os resultados que são esperados após a realização da coleta e análise dos dados, ou seja, quais resultados são esperados ao final da pesquisa)*

*O resultado esperado é uma analise de regressão em arquivo do R Studio que correlacione claramente as sensacões descritas pelos espectadores as notas altas e baixas, indicando quais palavras chave que descrevem essas sensações tendem a aumentar ou diminir a nota atribuida. Uma descrição pratica para que tomadores de decisão escolham algumas premissas para as suas produções e a distribuição e divulgação das mesmas.*

* **Cronograma de Atividades**

*(Adicionar as “Atividades planejadas”, assim como o período (tempo para desenvolver cada atividade) planejado para a realização de cada atividade, sendo que deverá ser adequado ao calendário de entregas das etapas do trabalho de conclusão de curso definido pela Coordenação. Marcar com um “x” a coluna que corresponde ao período planejado para desenvolver cada atividade planeja)*

*Existe um modelo pré organizado.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | | |
| **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| Analise de filmes do genero aventura | x |  |  |  |  |  |  |  |
| Definição das palavras chave | x |  |  |  |  |  |  |  |
| Projeto pesquisa | x |  |  |  |  |  |  |  |
| Título |  | x | x |  |  |  |  |  |
| Introdução |  | x | x |  |  |  |  |  |
| Objetivo |  | x | x |  |  |  |  |  |
| Materias e metodos |  | x | x | x |  |  |  |  |
| Resultados preliminares |  | x | x |  |  |  |  |  |
| Referencias bibliograficas e referencias |  | x | x | x |  |  |  |  |
| Cronograma de atividades | x |  |  |  |  |  |  |  |
| Entrega do TCC e o agendamento da defesa |  |  |  |  |  |  | x | x |
| O depósito da apresentação da defesa |  |  |  |  |  |  |  | x |

Projeto de Pesquisa; Resultados Preliminares; Entrega do Trabalho de Conclusão de Curso; Entrega da Apresentação da Defesa

* **Referências Bibliográficas**

*(Listagem das bibliografias citadas no projeto de pesquisa, seguindo rigorosamente as Normas do MBA USP ESALQ – Consulte o manual de “Normas para Elaboração do Trabalho de Conclusão de Curso” disponível no Sistema TCC)*

* *"Understanding Movie Ratings: An Empirical Study" por David B. Nieborg e Stefan Werning - 2010, Amsterdam, Holanda.*
* *"Movie Rating Prediction Based on Reviews Using Sentiment Analysis and Text Mining" por Shafiq R. Joty e Ehsan Hoque - 2016, Tóquio, Japão.*
* *"Mining User Generated Content for Aspect Based Movie Recommendation" por R. P. Sreeja e S. Sureshkumar - 2017, Tamil Nadu, Índia.*
* *"Predicting Movie Box Office Revenue Using Machine Learning Techniques" por Weiliang Zhao e Xiaohan Yin - 2017, Tianjin, China.*
* *"A Comparison of Text Mining Techniques in the Context of Movie Review Analysis" por Meryem El Baz, Noureddine Chenfour e Moulay Driss Rahmani - 2018, Marrakech, Marrocos.*
* *"Predicting Movie Success: A Neural Network Approach" por Brian Carrico e Ron VanDerScheer - 2018, Las Vegas, EUA.*
* *"On the predictability of movies' success: a data-driven approach" por Joana Silva, Sara Machado e Luís Sarmento - 2018, Nice, França.*
* *"A Hybrid Approach for Movie Success Prediction" por Ali Taheri e Somayyeh Mohammadynejad - 2019, Kuala Lumpur, Malásia.*
* *"Predicting Box-Office Success of Motion Pictures with Text Mining Techniques" por Mevlana Buldu e Hakan Ceylan - 2019, Ancara, Turquia.*
* *"Box Office Prediction of Movie Success using Machine Learning Techniques" por Jin Guo, Yongtao Hu e Yifan Wu - 2020, Nanjing, China.*
* *"Mineração de opiniões em avaliações de filmes: Uma análise comparativa de algoritmos de aprendizado de máquina" por Raphael Oliveira Guimarães e Adriano Veloso - publicado na Revista de Informática Teórica e Aplicada (RITA) em 2017, na cidade de Porto Alegre, Rio Grande do Sul, Brasil.*
* *"Análise de sentimentos em avaliações de filmes no IMDB" por William Menezes de Sousa e Anisio Franco - publicado no Congresso Brasileiro de Informática e Comunicação (CBIC) em 2017, na cidade de Natal, Rio Grande do Norte, Brasil.*
* *"Análise de sentimentos em avaliações de filmes: Uma abordagem utilizando o algoritmo Random Forest" por João Victor Gonçalves Barbosa e Guilherme Magalhães Ramos - publicado na Revista de Sistemas e Computação (RESCOM) em 2019, na cidade de Niterói, Rio de Janeiro, Brasil.*
* *"Análise de sentimentos em resenhas de filmes utilizando o algoritmo Naive Bayes" por Bruna Teixeira Mota e Flávio Henrique Teles Vieira - publicado na Revista Tecnologia e Sociedade em 2017, na cidade de Uberlândia, Minas Gerais, Brasil.*
* *"Mineração de Opiniões: Uma Análise de Sentimentos sobre Filmes Brasileiros no Twitter" por Pedro de Oliveira e Silva e Janaína Giraldi - publicado na Revista Brasileira de Computação Aplicada em 2017, na cidade de São Paulo, Brasil.*
* *"Análise de críticas de filmes em redes sociais utilizando Mineração de Texto" por Thiago Dias Martins e José Antonio Fernandes de Macedo - publicado no XVII Simpósio Brasileiro de Informática na Educação (SBIE), em 2006, na cidade de Araraquara, São Paulo, Brasil.*
* *"Análise de Sentimentos em Críticas de Filmes por meio de Algoritmos Clássicos de Aprendizado de Máquina" por Ana Júlia de Oliveira Bellini- São José dos Campos - Universidade Federal de São Paulo – UNIFESP Instituto de Ciência e Tecnologia Bacharelado em Ciência da Computação.*