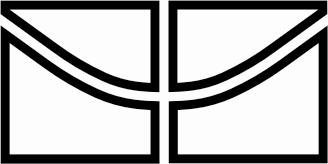
**Universidade de Brasília**



**Decanato de Pesquisa e Pós-Graduação**

Programa de Iniciação Científica – ProIC/UnB

**Instruções de preenchimento do Plano de Trabalho**

Na página 1, conforme o modelo insira os seguintes dados:

* Informe a área do conhecimento (Exatas, Humanas, Vida)
* Título do Projeto;
* Nome do Orientador;
* Unidade Acadêmica/Departamento do orientador
* Título do Plano de Trabalho;
* Nome do Aluno;
* Matrícula do Aluno;

No plano de trabalho preencha os tópicos:

**Introdução ao Plano de Trabalho**

Descrever nesta parte do plano de trabalho, os objetivos específicos (como parte dos objetivos gerais do projeto), o estado da arte, etc. Dois ou mais alunos não podem ter os mesmos objetivos específicos. A orientação é INDIVIDUAL.

**Metodologia do Plano de Trabalho**

Descrever a metodologia específica para este plano de trabalho. Deve-se evitar que dois alunos tenham metodologias semelhantes. A orientação é INDIVIDUAL.

**Resultados Esperados na Execução do Plano de Trabalho**

Descrever os resultados esperados na execução deste plano de trabalho.

**Etapas e Cronograma de Execução do Plano de Trabalho**

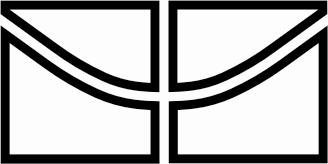
Liste as etapas, inclua ou elimine etapas se necessário. No cronograma, utilize os símbolos para marcar as etapas, apagando aqueles de maneira apropriada.

**Referências Bibliográficas**

Liste as referências bibliográficas.

Campus Universitário Darcy Ribeiro Pavilhão Anísio Teixeira Sala AT 012 - Brasília – DF CEP 70910-900

Telefones: (61) 3107.0822 - 3107.0824 – 3107.0825 – 3107-0826 e-mail: [pibic@unb.br](mailto:pibic@unb.br) <http://www.unb.br/administracao/decanatos/dpp/dific.html>

**Universidade de Brasília**



**Decanato de Pesquisa e Pós-Graduação**

Programa de Iniciação Científica – ProIC/UnB

ÁREA DO CONHECIMENTO**:** ( x ) EXATAS ( ) HUMANAS ( ) VIDA

Título do Projeto

**Aplicabilidade de Sistemas de Inteligência Artificial para Dados Abertos**

Orientador:

 **Marcelo Ladeira**

Unidade Acadêmica/Departamento:

 **Departamento de Ciência da Computação**

**PLANO de TRABALHO**



**Edital 2016 ProIC/CNPq/UnB**



Título do Plano de Trabalho

**Análise do Desempenho de Alunos da UnB usando um Sistema de Inteligência Artificial**

Aluno

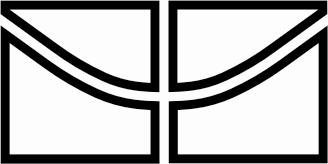
 **Gabriel Ferreira Silva**

Matrícula

 **14/0140131**

Campus Universitário Darcy Ribeiro Pavilhão Anísio Teixeira Sala AT 012 - Brasília – DF CEP 70910-900

Telefones: (61) 3107.0822 - 3107.0824 – 3107.0825 – 3107-0826 e-mail: [pibic@unb.br](mailto:pibic@unb.br) <http://www.unb.br/administracao/decanatos/dpp/dific.html>

**Universidade de Brasília**

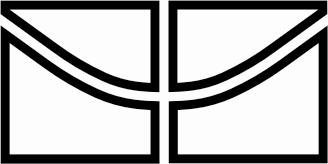


**Decanato de Pesquisa e Pós-Graduação**

Programa de Iniciação Científica – ProIC/UnB

Campus Universitário Darcy Ribeiro Pavilhão Anísio Teixeira Sala AT 012 - Brasília – DF CEP 70910-900

Telefones: (61) 3107.0822 - 3107.0824 – 3107.0825 – 3107-0826 e-mail: [pibic@unb.br](mailto:pibic@unb.br) <http://www.unb.br/administracao/decanatos/dpp/dific.html>

**Universidade de Brasília**



**Decanato de Pesquisa e Pós-Graduação**

Programa de Iniciação Científica – ProIC/UnB

**Plano de Trabalho (máximo de 5 páginas)**

**1. Introdução ao Plano de Trabalho**

Diversas instituições disponibilizam publicamente suas informações, seja por questões de transparência ou por perceberem que o estudo sobre tais dados pode indicar oportunidades a serem capitalizadas. Uma aplicação de interesse para dados abertos é, a partir de dados já conhecidos, tentar extrair informações que expliquem a relação entre os dados e, com base em tais informações, construir modelos que façam projeções sobre o futuro ou sobre dados que ainda não dispomos.

Seguindo esse premissa, este projeto visa a construção de um software capaz de fazer projeções acerca do desempenho futuro de um aluno na UnB, e possivelmente de sua remuneração no mercado após sair da faculdade. Isso só será possível pois foi conseguida com a UnB a disponibilização de dados de seus alunos. Tais dados incluem dados de perfil, o desempenho do aluno em cada matéria, se um aluno conseguiu ou não concluir a universidade, se houve participação em estágios ou PIBIC. Caso ele tenha ido ao mercado de trabalho privado, a sua remuneração inicial estará disponível.

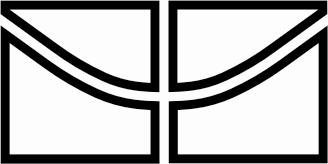
O sistema computacional descrito poderia ser usado para identificar quais alunos estão com maior risco de não conseguirem completar a universidade ou quais disciplinas oferecem mais risco ou quais disciplinas influenciam mais as outras. O software a ser desenvolvido seria baseado em técnicas de um campo na intersecção de Inteligência Artificial com Mineração de Dados: Aprendizagem de Máquina.

A área de Inteligência Artificial estuda a construção de sistemas computacionais inteligentes [1]. Já a área de Mineração de Dados estuda técnicas para a descoberta de padrões a partir de um Banco de Dados [2]. Na intersecção dessas duas áreas há um campo que vem despertando muito interesse: Aprendizagem de Máquina. O objetivo é conseguir extrair um padrão a partir de uma grande quantidade de dados, e usar tal padrão para com sucesso tentar prever o que ocorrerá com um sistema no futuro ou qual a maneira mais adequada de lidar com dados futuros.

O desenvolvimento de aprendizagem de máquina vem sendo bastante célere nos últimos tempos. Além disso, vem impactando áreas bastante diversas, como jogos [3], segurança digital [4] e sistemas de recomendação [5]. Em geral, os problemas estudados por Aprendizagem de Máquina se dividem em duas áreas: classificação e regressão [6]. Para a classificação, o resultado final assume um valor discreto. Técnicas típicas incluem Máquinas de Suporte Vetorial (SVM do inglês Support Vector Machine) e Árvores de Decisão. Para a regressão, o

Campus Universitário Darcy Ribeiro Pavilhão Anísio Teixeira Sala AT 012 - Brasília – DF CEP 70910-900

Telefones: (61) 3107.0822 - 3107.0824 – 3107.0825 – 3107-0826 e-mail: [pibic@unb.br](mailto:pibic@unb.br) <http://www.unb.br/administracao/decanatos/dpp/dific.html>

**Universidade de Brasília**



**Decanato de Pesquisa e Pós-Graduação**

Programa de Iniciação Científica – ProIC/UnB

resultado final assume um valor contínuo. Técnicas típicas incluem Regressão Linear e Redes Neurais [6].

**2. Metodologia do Plano de Trabalho**

Antes do desenvolvimento, uma etapa de pré-processamento será realizada com os dados, para garantir que eles estejam em um estado consistente. Em seguida, como é comum na ciência, uma análise exploratória dos dados será feita. Também como etapa inicial, um estudo acerca do estado da arte das técnicas de aprendizagem de máquina será feito.

Durante o desenvolvimento do modelo, utilizar-se-ão várias técnicas de Aprendizagem de Máquina. A escolha das técnicas adequadas será feita com base em sua adequabilidade a situação e a quantidade de dados que se dispõe. Para cada técnica, os dados serão separados em dois conjuntos disjuntos: um de

treino (para que o modelo seja capaz de aprender) e outro de teste (para avaliar quão bem o modelo aprendeu). A existência de *overfitting* será analisada

comparando a diferença de performance para os dados de treino e de teste. O desempenho das técnicas também será estimado, considerando para isso sua performance nos dados de teste.

**3. Resultados Esperados na Execução do Plano de Trabalho**

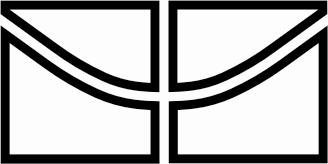
Espera-se que ao final do trabalho, um sistema inteligente com bom desempenho previsor tenha sido desenvolvido. Embora a capacidade de previsão do software seja um fator subjetivo, os desempenhos medidos no conjunto de teste serão fornecidos.

Possibilidades de atividades a serem previstas pelo sistema incluem: se um aluno irá formar ou não, o desempenho em uma matéria que ele ainda não fez, o IRA com o qual ele acabará a universidade, entre outras. Saber quais os alunos estão com maior risco de não concluir a universidade seria útil para a UnB, pois assim um maior direcionamento poderia ser dado a tal grupo. Já prever um desempenho de um aluno em uma matéria seria útil para o estudante ter, de modo personalizado, uma estimativa de quais matérias serão mais difíceis (permitindo assim um melhor planejamento para o decorrer do semestre). Por fim, uma estimativa do IRA com o qual o estudante terminará a universidade também poderia auxiliar no planejamento dele para ingressar no mercado de trabalho.

O software desenvolvido terá código aberto e será flexível e bem documentado, permitindo assim que o seu funcionamento externo seja entendido com clareza por qualquer um e que seu funcionamento interno seja compreendido por aqueles que conhecem Aprendizagem de Máquina. Pela maneira como o software será desenvolvido, o software deve poder ser adaptado ou tomado como base para trabalhos com temas semelhantes.

Campus Universitário Darcy Ribeiro Pavilhão Anísio Teixeira Sala AT 012 - Brasília – DF CEP 70910-900

Telefones: (61) 3107.0822 - 3107.0824 – 3107.0825 – 3107-0826 e-mail: [pibic@unb.br](mailto:pibic@unb.br) <http://www.unb.br/administracao/decanatos/dpp/dific.html>

**Universidade de Brasília**



**Decanato de Pesquisa e Pós-Graduação**

Programa de Iniciação Científica – ProIC/UnB

**4. Etapas e Cronograma de Execução do Plano de Trabalho**

Etapa 1 – Detalhamento do Plano de Trabalho. Busca por outras possíveis aplicações e metodologias adequadas.

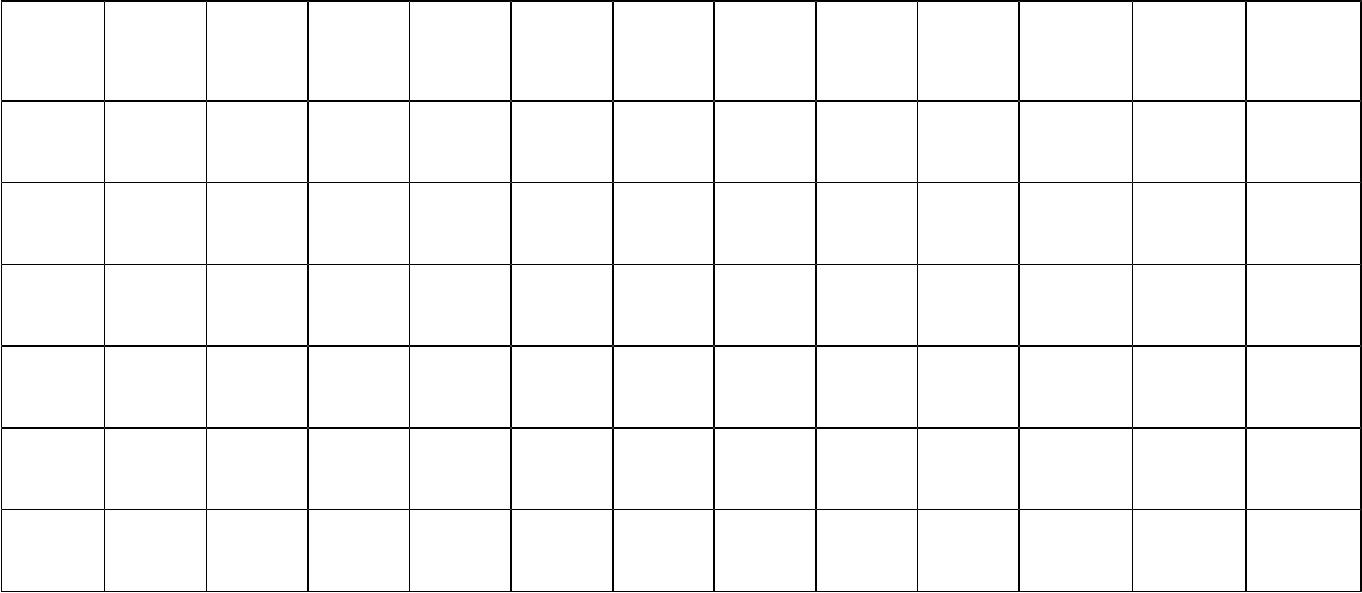
Etapa 2 – Estudo de Aprendizagem de Máquina com pesquisa em livros, artigos e MOOC's. As principais técnicas serão estudadas, assim como seu contexto de aplicação. Técnicas específicas para o contexto do trabalho serão procuradas.

Etapa 3 – Pré-processamento dos dados, de modo a identificar dados incompletos ou outros problemas. Decidir como proceder para o caso de dados inconsistentes.

Etapa 4 – Análise Exploratória dos dados, busca por informações interessantes nos dados.

Etapa 5 – Aplicação do conhecimento adquirido na etapa 2 através do desenvolvimento do modelo. Análise do desempenho via performance nos dados de teste e identificação de overfitting via disparidades na performance de treino e teste.

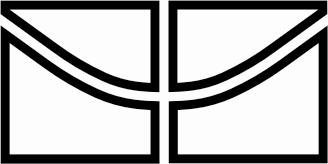
Etapa 6 – Preparar o relatório final e o pôster para apresentação no PIBIC



|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Etapa | Mês | Mês | Mês | Mês | Mês | Mês | Mês | Mês | Mês | Mês | Mês | Mês |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

Campus Universitário Darcy Ribeiro Pavilhão Anísio Teixeira Sala AT 012 - Brasília – DF CEP 70910-900

Telefones: (61) 3107.0822 - 3107.0824 – 3107.0825 – 3107-0826 e-mail: [pibic@unb.br](mailto:pibic@unb.br) <http://www.unb.br/administracao/decanatos/dpp/dific.html>

**Universidade de Brasília**



**Decanato de Pesquisa e Pós-Graduação**

Programa de Iniciação Científica – ProIC/UnB

**5. Referências Bibliográficas**

[1] - Russell, Stuart, Peter Norvig, and Artificial Intelligence. "A modern approach."*Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs* 25 (1995):

27.

1. - Aggarwal, Charu C.*Data mining: The textbook*. Springer, 2015.
2. - Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search."*Nature* 529.7587 (2016): 484-489.
3. - https://people.csail.mit.edu/kalyan/AI2\_Paper.pdf
4. - Koren, Yehuda. "The bellkor solution to the netflix grand prize."*Netflix* *prize documentation* 81 (2009).
5. – Abu-Mostafa, Yaser S., Malik Magdon-Ismail, and Hsuan-Tien Lin. *Learning from data*. Berlin, Germany: AMLBook, 2012.

Campus Universitário Darcy Ribeiro Pavilhão Anísio Teixeira Sala AT 012 - Brasília – DF CEP 70910-900

Telefones: (61) 3107.0822 - 3107.0824 – 3107.0825 – 3107-0826 e-mail: [pibic@unb.br](mailto:pibic@unb.br) <http://www.unb.br/administracao/decanatos/dpp/dific.html>