**Componente:** Aprendizado de Máquina I

**Aula: AULA 5**

**TRILHA DE APRENDIZAGEM**

1. **TÍTULO DA AULA (RECURSO OBRIGATÓRIO)**

|  |
| --- |
| 5. Árvores de Decisão e Seleção de Atributos |

1. **OBJETIVO DA AULA (RECURSO OBRIGATÓRIO)**

|  |
| --- |
| Nesta aula você vai aprender:   * Sobre Classificador de Árvores de Decisão * O que é Ganho de Informação ou Informação Mútua * Como empregar esses e outros conceitos na Seleção de Atributos para o Aprendizado |

1. **ORDEM DOS RECURSOS NA TRILHA**

Na primeira coluna, constam os recursos que você poderá indicar no formulário. Preencha na segunda coluna a ordem que você deseja que os recursos fiquem na trilha de aprendizagem, numerando-os a partir de 1.

Considere como quantidade ideal de recursos na Trilha entre 5 e 6 itens.

|  |  |
| --- | --- |
| **Recursos** | **Ordem** |
| Ponto de Partida | Este recurso não pode ter a ordem alterada |
| Texto de Apoio (obrigatório) | 1 |
| Professor Resolve 1 (opcional) | 2 |
| Praticando 1 (opcional) |  |
| Professor Resolve 2 (opcional) |  |
| Praticando 2 (opcional) |  |
| Momento com o Professor (opcional) | 4 |
| Explorando Conceitos (opcional) | 3 |
| Aplicando Conhecimento | Este recurso não pode ter a ordem alterada |

1. **PONTO DE PARTIDA (RECURSO OBRIGATÓRIO)**

Toda aula iniciará com uma videoaulachamada “Ponto de Partida”.  **Você poderá gravar diversos vídeos ao decorrer da trilha, desde que totalizem 20 minutos por aula**.

Em decorrência da pandemia de Covid-19, você poderá optar por gravar as videoaulas no estúdio, mediante prévio agendamento, ou no formato “vídeo caseiro”, utilizando seus próprios recursos. Indique, a seguir, o formato de gravação:

**☐ Gostaria de gravar a aula no estúdio.**

**X Gostaria de gravar a aula no formato homevídeo.**

Preencha a seguir as informações necessárias sobre a videoaula “Ponto de Partida”:

|  |
| --- |
| **PONTO DE PARTIDA** |
| Duração da videoaula: (Sugestão: de 5 a 8 minutos)  Título da videoaula: Árvores de Decisão e o Método de Particionamento Recursivo  Descrição da videoaula: Vamos aprender aqui como funciona uma Árvore de Decisão para Classificação? É isso que eu vou mostrar para você nesta vídeo-aula. Aprenda, também, intuitivamente o que é o ganho de informação!  Recursos que serão utilizados: Browser com acesso à Internet e Google Colaboratory. |

1. **TEXTO DE APOIO (RECURSO OBRIGATÓRIO)**

Por uma questão de praticidade, este item está posicionado no final deste documento.

1. **PROFESSOR RESOLVE + PRATICANDO (ESTA ATIVIDADE PODE GERAR NOTA DE PARTICIPAÇÃO)**

Neste recurso, o aluno terá acesso a um combinado de videoaula(s) e exercícios para praticar. Você poderá escolher como esses recursos serão apresentados na trilha.

Escolha uma das opções a seguir:

**FORMATO 1 (PROFESSOR RESOLVE + PRATICANDO)** – O aluno assistirá à(s) videoaula(s) e, em seguida, realizará uma atividade. Nesse caso, a atividade deve ser, preferencialmente, de autocorreção. Se não for de autocorreção, você deverá elaborar um gabarito para divulgação.

**FORMATO 2 (PRATICANDO + PROFESSOR RESOLVE)** – Neste caso, o aluno realiza a atividade e, em seguida, assiste à(s) videoaula(s) com a resolução dos exercícios.

**FORMATO 3 (PROFESSOR RESOLVE)** – Neste caso, o aluno assistirá à(s) videoaula(s) e não haverá exercício vinculado.

Quantidade de videoaulas que você pretende gravar **neste recurso**: Escolher um item.

Você poderá optar por gravar as videoaulas no estúdio, mediante prévio agendamento, ou no formato “vídeo caseiro”, utilizando seus próprios recursos. Indique, a seguir, o formato de gravação:

**Gostaria de gravar a aula no estúdio.**

**Gostaria de gravar a aula no formato homevídeo.**

Preencha o quadro abaixo para cada videoaula que você pretende gravar neste recurso.

|  |
| --- |
| **PROFESSOR RESOLVE** |
| Duração da videoaula: 5-8min  Título da videoaula: Aplicando o Modelo de Árvore de Decisão e Calculando o Ganho de Informação  Descrição da videoaula: Neste vídeo vamos entender como aplicar os estimador de Árvores de Decisão do SciKit-Learn e como obter o ganho de informação de atributos!  Recursos que serão utilizados: Navegador Web com acesso à Internet e Colab. |

|  |
| --- |
| **PRATICANDO**  **Faça aqui a descrição da atividade** |
| Descrição da atividade: (Preencha aqui a atividade proposta. Caso você opte pelo Formato 1, poderá utilizar exercícios fechados, como questões de múltipla escolha, assim, o aluno receberá o feedback automaticamente. Se a atividade não for de autocorreção, lembre-se de incluir o gabarito com a resolução, o qual será divulgado pelo professor mediador). |

1. **MOMENTO COM O PROFESSOR**

Ao decorrer do componente, os alunos podem ter contato direto com o professor mediador, em três formatos diferentes:

* FORMATO 1 (BBB) – O professor conteudista prevê uma aula “ao vivo” na trilha. Não se trata de plantão de dúvidas. É um encontro com propósito: pode ser a explicação de um conceito, ou a resolução de um exercício, ou uma discussão a partir de um vídeo do YouTube a que todos assistiram. Metade das aulas de seu curso deve ter um encontro neste formato. Você indicará a temática para esse BBB, e o professor mediador será o responsável por planejar e executar essa aula.
* FORMATO 2 (FÓRUM DE DISCUSSÃO) – O professor conteudista apresenta uma questão problematizadora e propõe uma discussão com a turma. Você pode oferecer diversos recursos para essa discussão: um objeto de aprendizagem, um texto, um podcast, um link para site ou vídeo. O professor mediador, então, acompanhará o Fórum ao decorrer da disciplina, incentivando a interação entre os alunos e contribuindo com as postagens.
* FORMATO 3 (FÓRUM DE DÚVIDAS) – O professor conteudista pode reservar um espaço para que os alunos postem suas dúvidas. O professor mediador as responderá, de forma que todos possam ter acesso ao conteúdo.

**Lembre-se de que metade das aulas deve ter uma videoconferência, e você poderá escolher em quais aulas o BBB será realizado.**

Para a aula que está elaborando, escolha uma das opções a seguir:

**Gostaria de aplicar o BBB nesta aula.**

**Gostaria de aplicar um Fórum de Discussão nesta aula.**

**Gostaria de aplicar um Fórum de dúvidas nesta aula.**

**Não vou usar nenhum recurso “Momento com o professor nesta aula”.**

Agora, em função da escolha acima, preencha o quadro correspondente:

Se você optou por propor uma aula via BBB, indique a temática da aula e seu objetivo:

|  |
| --- |
| **MOMENTO COM O PROFESSOR – BBB** |
| Assunto que será tratado na aula: Árvores de Decisão e Ganho de Informação  Orientações para o mediador: Divida a sessão em três temas principais.   1. Árvores de Decisão e sua aplicação com o scikit-learn.(15min). Reforce aqui, novamente a semelhança na construção da solução com o uso de outros estimadores. Destaque a diferença do uso de Àrvores de Decisão: elas empregam a probabilidade dos valores e não os valores e funções de distância como nos métodos anteriores. 2. O que é Entropia, Ganho de Informação e qual a sua importância para a seleção de atributos de modelos. (15min) Destaque que existem diferentes formas de se medir o Ganho de Informação, embora o conceito seja o mesmo. 3. A Seleção de Hiperparâmetros Versus a Seleção de Modelos (15min). Destaque as diferenças e prepare o aluno para o que verá na aula seguinte. Discuta a necessidade de automatizar essas tarefas, comentando sobre os recursos do Scikit-learn para isso e até mesmo a ideia de AutoML.   Reserve alguns minutos para dúvidas dos alunos ou discuta outros aspectos da seleção de features e engenharia de atributos.  Duração: (Indique a duração prevista para a realização desse BBB. Exemplo: 60 minutos) |

Se você optou por propor um fórum de discussão, indique a proposta no bloco a seguir:

|  |
| --- |
| **MOMENTO COM O PROFESSOR – FÓRUM DE DISCUSSÃO** |
| Descrição do fórum: Outros critérios de Seleção de Atributos.  Pesquise outros critérios que podem ser empregados para a seleção de atributos e compartilhe com seus colegas. Dê exemplos de aplicação dessa técnica.  Pontuação: 0,5  Recursos do fórum: Não deixe de indicar aqui referências as técnicas/critérios indicados. |

Se você optou por propor um fórum de dúvidas, não é necessário adicionar outras informações, pois já temos um formato específico para esse recurso.

1. **EXPLORANDO OS CONCEITOS**

Neste recurso, você terá a liberdade de propor materiais complementares para esta aula. Você poderá indicar animação, links externos, games, simuladores, exercícios resolvidos, podcast, entre outros.

**Gostaria de propor um material complementar nesta aula.**

**Não gostaria de propor um material complementar nesta aula.**

Se você optou por propor um material complementar, indique a proposta no bloco a seguir:

|  |
| --- |
| **EXPLORANDO OS CONCEITOS** |
| Descrição do recurso:   * Acesse https://ml-playground.com/# e experimente alguns modelos de classificação sem qualquer programação. Se estiver interessado ainda pode verificar como exportar e importar dados do https://ml-playground.com/# para o Python. Dica, os dados são salvos e carregados em formato .json. * O exemplo clássico Árvores de Decisão de Mitchell, Tom Michael (1997) em Machine Learning é também explorado em Sayad, Saed (2021). An Introduction to Data Science, Disponível em: https://www.saedsayad.com/data\_mining\_map.htm e ajuda a entender em um passo a passo como é construída uma árvore de decisão e são calculados o ganho de informação. * Acesse Jake VanderPlas. Python Data Science Handbook Disponível em: https://jakevdp.github.io/PythonDataScienceHandbook/ e veja o capítulo In-Depth: Decision Trees and Random Forests, em especial o exemplo: Example: Random Forest for Classifying Digits classificam dados de imagens que não exploramos aqui. |

1. **APLICANDO CONHECIMENTO (RECURSO OBRIGATÓRIO)**

Ao final de cada aula, o aluno realizará uma atividade de encerramento. Esta atividade é avaliativa, e você deverá indicar a pontuação dela no quadro de pontuação.

Indique a atividade no quadro a seguir:

|  |
| --- |
| **APLICANDO CONHECIMENTO** |
| Descrição da atividade:  Acesse o questionário no Moodle referente a Aula 5 e resolva os exercícios. Lá você encontrará questões conceituais e práticas sobre Árvores de Decisão, Ganho de Informação e Seleção de Atributos. A parte prática é bastante semelhante ao que tratamos no vídeo Professor Resolve anteriormente. Assista ou reveja o vídeo antes de realizar a tarefa. Para os exercícios práticos empregue o notebook:  <https://colab.research.google.com/github/Rogerio-mack/Machine-Learning-I/blob/main/ML5_DecisionTrees_ex.ipynb>  Depois de realizada a tarefa você terá acesso ao gabarito e a solução desses exercícios.  Pontuação: 1 |

1. **GABARITO – APLICANDO CONHECIMENTO (RECURSO OBRIGATÓRIO)**

Forneça o gabarito aqui ou, se for disponibilizar o gabarito de outra forma, converse com a Designer, para que ela lhe oriente.

|  |
| --- |
| **GABARITO** |
| Disponível em:  Moodle  <https://colab.research.google.com/github/Rogerio-mack/Machine-Learning-I/blob/main/ML5_DecisionTrees_ex_solucao.ipynb> |

**TABELA DE PONTUAÇÃO DA AULA**

Preencha a tabela abaixo com a pontuação aplicada nessa aula e a soma com as demais aulas.

**Lembrete:** a soma de todas as atividades do componente totaliza 10 pontos.

|  |  |
| --- | --- |
| Total de pontos aplicados nessa aula. | 1,5\_\_\_\_ /10,0 pontos |
| Total dos pontos que temos até o momento (soma das aulas anteriores). | 6\_\_\_\_ / 10,0 pontos |

1. **TEXTO DE APOIO (RECURSO OBRIGATÓRIO)**

Em todas as aulas, você deverá produzir um texto de apoio que apresente o conteúdo essencial à aula. O texto de apoio deve conter os aspectos mais importantes, como conceitos, ideias, teses, trechos etc., a respeito do assunto. **O texto deve ser de autoria própria, permitindo que o aluno aprofunde o conhecimento e assimile os conteúdos**.

Apresente consistência, ou seja, organize informações importantes ou sinalize caminhos relevantes para a construção do conhecimento.

**Orientações para a elaboração:**

- Para cada aula, você deve desenvolver cerca de **10 páginas (mínimo 8, máximo 12).** Utilize a seguinte formatação: fonte Arial, tamanho 12, espaçamento 1,5 entrelinhas e alinhamento justificado.

- Use sempre a linguagem dialógica, de forma que você passe para o aluno a ideia de que está conversando com ele. Exemplos de expressões: “Você reparou como...”, “Você já estudou XYZ, agora veja...”, “Você deve prestar atenção em XYZ...”, “Percebeu como fica tal coisa...”

- Indique imagens ilustrativas, assim como gráficos e esquemas. O CEDaD detém a assinatura do banco de imagens Getty Images, não gerando, assim, custo ao professor conteudista. Busque as imagens que deseja utilizar e indique o link exato no material: <https://www.gettyimages.com.br/>

- Ao longo do texto, utilize elementos extratextuais para interagir com o aluno. Exemplos: Saiba mais, Curiosidades, Atenção, entre outros. Para identificar esses elementos, utilize a seguinte formatação: (fonte em vermelho e entre parênteses). Exemplo:

*[...]*

*A classe Pilha estenderá a classe Vetor e, portanto, terá acesso aos métodos públicos dessa classe. É muito importante que você retome os conceitos de herança que você já estudou em componentes anteriores!*

- Veja um modelo desse texto no ambiente de entrega de materiais.

Escreva nas páginas a seguir o texto de apoio desta aula:

|  |
| --- |
| TÍTULO DO TEXTO DE APOIO: (Crie um título para este texto)  TEXTO: (Escreva nas páginas a seguir o texto de apoio desta aula.)  Documento postado em word à parte no Moodle. |