**Componente:** Análise Exploratória de Dados

**Aula: AULA 8**

**TRILHA DE APRENDIZAGEM**

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

|  |
| --- |
| Probabilidade Condicionada, Bayes e Informação Mútua |

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

|  |
| --- |
| Nesta aula você aprenderá:   * Como empregar com probabilidades condicionadas, além de revisitar alguns conceitos básicos de probabilidade * O que é e como empregar o Teorema de Bayes * Como obter e empregar a informação mútua para a seleção de atributos (\*feature selection\*) para Análise de Dados e para problemas de Aprendizado de Máquina |

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) |  |
| 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: Teorema de Bayes  Descrição da videoaula: Acompanhe essa vídeo-aula para aprender um pouco sobre probabilidades condicionadas e esse importante teorema que é o Teorema de Bayes. Você vai entender como probabilidades podem ser atualizadas à medida que adicionamos 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: Cálculo de Informação Mútua  Descrição da videoaula: **Na Análise Exploratória de Dados uma das análises mais importantes é determinar o que o valor de um atributo pode dizer sobre os valores de outro. Isso tem uma série de aplicações, em particular para a seleção de atributos para o aprendizado de máquina.** Acompanhe então essa vídeo-aula para aprender como podemos fazer isso em R ou Python!  Recursos que serão utilizados: Navegador Web, Google Colaboratory. |

|  |
| --- |
| **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).  **Informação Mútua: Feature Selection**  Na Análise Exploratória de Dados uma das análises mais importantes é determinar o que o valor de um atributo pode dizer sobre os valores de outro. Isso tem uma série de aplicações, em particular para a seleção de atributos para o aprendizado de máquina. Tratamos isso no nosso Professsor Resolve, e no material da aula você encontra como empregar a informação mútua para selecionar os atributos mais relevantes para determinar uma classe em R (Aula 8, Feature Selection).  Este exemplo <https://colab.research.google.com/github/Rogerio-mack/Analise_Exploratoria_de_Dados/blob/main/AED_8_Praticando.ipynb> mostra como fazer isso em Python.  **Atividade**  Empregue o exemplo para fazer uma seleção de atributos de classificação (atributos relevantes para a determinação de uma classe) de um conjunto de dados de interesse e compartilhe com os seus colegas no fórum.  Você ainda pode buscar por outros exemplos na internet de outras técnicas de seleção de features, incluindo seleção de features para regressão (em vez de classificação como apresentamos aqui). Neste caso não deixe explicar um pouco da técnica empregada e indicar as referências empregadas. ;-) |

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:  Orientações para o mediador:  Duração: |

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: O Datafólio do seu projeto é opcional. Mas você pode postá-lo aqui para compartilhar com os seus colegas e discutirmos a importância da comunicação nas análises que são efetuadas. Você pode também querer falar sobre a sua experiência em produzir um datafólio.  Pontuação: 0,5  Recursos do fórum: - |

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:  1. Acesse https://www.kaggle.com/ryanholbrook/mutual-information empregue o modelo de código fornecido para calcular os atributos que mais trazem informação sobre o preço dos veículos na base `Cars93` (Dica: crie a base Cars93 em Python para empregar as funções do exemplo em Python).  2. Acesse Sun, Dennis. \*\*Introduction to Probability\*\* em: https://dlsun.github.io/probability/ e leia o capítulo 9 para descobrir como interpretar geometricamente a Lei de Bayes.  3. As Probabilidades Condicionais e o pensamento Baysiano não são tão intuitivas como o pensamento frequentista de probabilidade. Acesse esse texto \*\*Monty Hall Problem\*\* em https://brilliant.org/wiki/monty-hall-problem/?quiz=monty-hall e conheça um caso clássico onde o pensamento bayesiano leva um um resultado bastante contra-intuitivo, embora seja algo que podemos observar na prática.  3. Esse é um tema bastante avançado, mas se você quiser saber sobre distribuições de probabilidade no modelo Baysiano, pode acessar essa introdução que ainda traz códigos em R: \*\*Bayesian models in R\*\* https://www.r-bloggers.com/2019/05/bayesian-models-in-r-2/ |
|  |

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 8 e resolva os exercícios. Os exercícios dessa aula envolvem questões teóricas e práticas que poderão ser resolvidas em R ou Python. Antes de iniciar a tarefa assista o Professor Resolve desta aula que traz dicas para o tema de informação mútua. O enunciado dos exercícios práticos pode ser acessado em  <https://github.com/Rogerio-mack/Analise_Exploratoria_de_Dados/blob/main/AED_8_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:  https://github.com/Rogerio-mack/Analise\_Exploratoria\_de\_Dados/blob/main/AED\_8\_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). | 10\_\_\_\_ / 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. |