**Componente:** Aprendizado de Máquina I

**Aula: AULA 2**

**TRILHA DE APRENDIZAGEM**

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

|  |
| --- |
| 6. Validação Cruzada e GridSearch |

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

|  |
| --- |
| Nesta aula você vai entender:   * O que são os conjuntos de Validação e Teste * Como e por que aplicar a Validação Cruzada dos modelos * Como empregar GridSearch do Scikit-Learn para selecionar os hiperparâmetros de um modelo |

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: Validação Cruzada  Descrição da videoaula: Nosso ponto de partida dessa aula é entendermos o por que e como funciona a validação cruzada. Esse é um ponto muito importante da avaliação da eficiência dos modelos e normalmente é a forma mais empregada.  Recursos que serão utilizados: Browser com acesso a Internet. 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: Selecionando Hiperparâmetros de um Modelo  Descrição da videoaula: Vamos entender aqui o que são os Hiperparâmetros de um modelo e como podemos empregar o GridSearch do scikit-learn para a seleção de melhores Hiperparâmetros.  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). |

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: (Preencha aqui a proposta do fórum. Você pode contextualizar o assunto e, em seguida, apresentar a questão problematizadora.)  Pontuação: (O fórum pode ser considerado um recurso avaliativo, assim, estimulamos o aluno a participar da discussão. Indique neste item o valor da nota. Lembre-se de que essa nota será somada à atividade final “Aplicando conhecimento” e, ao término do componente, deve totalizar 10 pontos. Se este fórum for apenas para tirar dúvidas, escreva aqui “FÓRUM DE DÚVIDAS”.)  Recursos do fórum: (Indique aqui os possíveis recursos que você utilizará neste fórum –texto, podcast, objeto de aprendizagem, link etc.) |

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:   * A documentação do scikit-learn é uma excelente fonte de consulta para todos as técnicas que empregamos aqui. Mas ela é também bastante extensa e selecionamos aqui alguns dos links principais que podem interessar:   + Cross-validation: evaluating estimator performance, Disponível em: https://scikit-learn.org/stable/modules/cross\_validation.html#cross-validation   + Tuning the hyper-parameters of an estimator, Disponível em: https://scikit-learn.org/stable/modules/grid\_search.html#grid-search, e aqui você pode buscar mais referência de como empregar a seleção de hiperparâmetros com múltiplas métricas.   + Parameter estimation using grid search with cross-validation, Disponível em: <https://scikit-learn.org/stable/auto_examples/model_selection/plot_grid_search_digits.html#sphx-glr-auto-examples-model-selection-plot-grid-search-digits-py> * Mas para saber mais sobre esses procedimentos o melhor mesmo é praticar. Você pode empregar então o último código dessa trilha, que é um modelo final de código para seleção de hiperparâmetros de um classificador, para fazer o mesmo para o classificador logístico, um regressor linear ou ainda empregar um outro conjunto de dados de seu interesse. Essas são atividades que, de fato, irão ajudá-lo a saber mais sobre o que tratamos aqui. * Nesta aula fizemos a normalização dos dados de entrada antes de aplicarmos os estimadores. Vamos saber mais sobre isso? De um modo simples, dados são normalizados ou redimensionados para trazer todas as variáveis em proporção umas com as outras. Existem diferentes formas de normalização e redimensionamento de dados e aqui aplicamos apenas uma delas. Acesse o artigo When to perform a Feature Scaling? de Raghav Vashisht, disponível em: https://www.atoti.io/when-to-perform-a-feature-scaling/ para saber mais sobre isso e veja também Compare the effect of different scalers on data with outliers disponível em: https://scikit-learn.org/stable/auto\_examples/preprocessing/plot\_all\_scaling.html para conhecer outros estimadores de normalização do scikit-learn. |

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 6 e resolva os exercícios. As questões estão divididas em uma parte conceitual e uma parte prática em que você irá desenvolver um código de solução bastante semelhante ao que tratamos no vídeo Professor Resolve. Para os exercícios prático empregue o notebook:  <https://colab.research.google.com/github/Rogerio-mack/Machine-Learning-I/blob/main/ML6_CV_GridSearch_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 e em  <https://colab.research.google.com/github/Rogerio-mack/Machine-Learning-I/blob/main/ML6_CV_GridSearch_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\_\_\_\_ /10,0 pontos |
| Total dos pontos que temos até o momento (soma das aulas anteriores). | 7\_\_\_\_ / 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. |