**Componente:** Aprendizado de Máquina I

**Aula: AULA 2**

**TRILHA DE APRENDIZAGEM**

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

|  |
| --- |
| 7. Seleção de Modelos |

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

|  |
| --- |
| Nesta aula você vai aprender sobre:   * Como selecionar melhores modelos de Classificação e Regressão * Sobre Outros Estimadores importantes * Como empregar modelos de aprendizado supervisionado para a Classificação de Imagens de Dígitos e Séries Temporais |

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: Por que e Como Selecionar diferentes modelos?  Descrição da videoaula: Vamos discutir nesta vídeo aula qual a importância de empregarmos diferentes modelos e como podemos fazer a seleção entre modelos que competem.  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: Na prática: vamos criar modelos diferentes e selecionar um melhor modelos dentre eles?  Descrição da videoaula: Vamos criar aqui alguns modelos diferentes para um problema de classificação e observar as métricas do scikit-learn para a seleção de um melhor modelo.  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: Da Seleção de Modelos à Escolha de Ferramentas  Orientações para o mediador: Divida a sessão em 4 temas principais.   1. Discuta a Seleção de Modelos (10min). Paradigmas diferentes e a importância de métricas; Enfatize que nenhum modelo é melhor a priori; Provoque uma discussão: *Modelos de Deep Learning são melhores que outros modelos?* 2. Discuta o uso de Python x R (10min). Destaque os nichos de um e outro. R: bioinformática, ciências sociais, modelos estatísticos (séries temporais), econometria; Python: Aprendizado de Máquina em geral; Modelos de Deep Learning; dados Multimídia. 3. Discuta o uso de outras Ferramentas (10min). Necessidade de Pacotes (scikit-learn, caret etc.); Limitação das linguagens e necessidade de um ecossistema (big data) para processamento de grandes volumes de dados; Limitações para processamento realtime; etc. 4. Reserve ao menos (15min) da aula para apresentar a tarefa do projeto, seu template para tirar dúvidas dos alunos.   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: (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:   * Interessado em saber mais sobre o uso de Aprendizado de Máquina para previsão de Séries Temporais? Acesse Kaggle Time Series Um curso online com uma aborgem de Aprendizado de Máquina e Engenharia de Features aplicado à Séries Uni e Multivariadas <https://www.kaggle.com/learn/time-series>. * Modelos Estatísticos para Séries Temporais. Previsões de Séries Temporais são feitas à muito tempo, mesmo antes de surgirem modelos de aprendizado de máquina. Aqui você pode explorar os modelos estatísticos tradicionais de Séries Temporais aplicados com Python. Time Series Forecasting in Python | Book Online & Videos https://livebook.manning.com/book/time-series-forecasting-in-python-book/welcome/v-8/ <https://www.youtube.com/channel/UC-0lpiwlftqwC7znCcF83qg/featured> * Acesse o Site de THE MNIST DATABASE of handwritten digits http://yann.lecun.com/exdb/mnist/ para ver alguns dos benchmarks de classificação da base MNIST. Yann LeCun é um dos mais influentes cientistas de IA e Aprendizado de Máquina, um dos criadores dos modelos de Deep Learning e Cientista Chefe de IA da Meta (FaceBook). * Explore diferentes modelos de classificação com o scikit-learn. Você pode começar por aqui:   + Choosing the right estimator https://scikit-learn.org/stable/tutorial/machine\_learning\_map/index.html, ou   + Classifier comparison <https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html> * Que tal criar seus próprios projetos empregando os mesmos dados que empregamos aqui mas com outros modelos, ou empregue outros dados de interesse. Procure explorar modelos mais elaborados como SVC (Support Vector Machines) e AdaBoostClassifier, e empregar dados com mais classes de saída e maior volume de casos. Você também pode empregar modelos Árvore ou mesmo de K-vizinhos mais próximos para regressões não lineares. Os modelos código e o que você aprendeu até aqui são bastante suficientes para você avançar bastante rapidamente nesses projetos. |

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:  Objetivo  Neste trabalho você deve obrigatoriamente trabalhar em grupo de 2-4 alunos e desenvolver uma solução de aprendizado supervisionado para algum problema de interesse do grupo. Os problemas poderão ser de classificação ou de regressão.  Descrição do Projeto  O trabalho deve ser original e desenvolvido em R ou Python notebook com um conjunto de dados próprios e originais definidos. Entendam como dados próprios e originais dados que são originais de empresas ou entidades, podendo ser bases públicas. Dados de ONGs, empresas privadas ou governo, são exemplos de bases válidas. Bases de repositórios públicos como Kaggle e UCI, que fornecem bases prontas não são originais. Tenha em mente: você deve construir uma base de dados nova para o seu problema, é essa a ideia de dados originais. Se tiver dúvidas consulte o professor antes de empregar uma base.  Os dados devem suportar um problema relevante e adequado para o aprendizado supervisionado. Assim, não podem ser empregadas bases artificiais e problemas muito simples. Devem ser evitadas também Séries Temporais como séries financeiras (como ações) e de dados climáticos que requerem outras técnicas não detalhadas aqui.  As bases devem conter um mínimo de 1000 instâncias. Consulte o professor para uso de um número menor de instâncias para dados próprios de uma empresa por exemplo.  Você deve implementar ao menos dois modelos, comparar e discutir os seus resultados. Seu notebook deve ser 100% executável e incluir todo necessário para construção dos dados, entendimento do problema e da solução.  Deve ainda ser entregue uma apresentação em \*\*vídeo (até 5min)\*\* dos resultados.  Acesse para um template do projeto:  <https://github.com/Rogerio-mack/Analise_Exploratoria_de_Dados/blob/main/ML7_ex_projeto.ipynb>  para mais detalhes da atividade, como as bases de dados elegíveis e os templates.  Pontuação: 2 |

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/ML3\_ ML7\_SelecaoDeModelos\_ex\_solucao.ipynb](https://colab.research.google.com/github/Rogerio-mack/Machine-Learning-I/blob/main/ML3_%20ML7_SelecaoDeModelos_ex_solucao.ipynb%20) |

**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). | 2\_\_\_\_ / 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. |