



Plano de Aprendizagem

1 Código e nome da disciplina

ARA0168 TÓPICOS DE BIG DATA EM PYTHON

2 Natureza

Extensão

3 Carga horária semestral

80

4 Carga horária semanal

4 horas de Extensão

5 Perfil docente

O docente deve ser graduado em Ciência da Computação, Sistemas de Informação, Engenharia de Computação, ou afins. É desejável a Pós-Graduação Stricto Sensu (Mestrado e/ou Doutorado) na área do curso ou áreas afins.

É desejável que o docente possua experiência profissional na área de Big Data e com a linguagem Python além de conhecimentos e habilidades teórico-práticos, capacidade de comunicação, interação e fluência digital para utilizar ferramentas necessárias ao desenvolvimento do processo de ensino-aprendizagem (SGC, SAVA, BdQ e SIA). Importante, também, o conhecimento do Projeto Pedagógico dos Cursos que a disciplina faz parte na Matriz Curricular.

É necessário que o docente domine as metodologias ativas inerentes à educação por competências, em especial a aprendizagem baseada em projetos e ferramentas digitais que tornem o processo mais interativo. A articulação entre ensino, pesquisa e extensão deve ser o eixo direcionador das estratégias utilizadas pelo docente. Além disto, é imprescindível que o docente estimule o autoconhecimento e autoaprendizagem entre seus alunos.

6 Área temática

Em atendimento à Resolução CNE/CES nº 7, de 18 de dezembro de 2018, que Estabelece as Diretrizes para a Extensão na Educação Superior Brasileira e dá outras providências, a área temática priorizada neste plano é tecnologia, produção e trabalho.

7 Linha eixo de extensão e pesquisa

O projeto extensionista desta disciplina está, em primeira análise, ligada ao eixo Científico-Acadêmico. Contudo, considerando os benefícios trazidos às partes interessadas, o projeto também está alinhado com os eixos de Empregabilidade, Empreendedorismo e Inovação; e Sustentabilidade.

8 Competências a serem trabalhadas

Com base na proposta institucional para a formação do egresso e as competências gerais e específicas desenvolvidas no curso, previstas em seu PPC, e em consonância com a Resolução CNE/CES nº 7, de 18 de dezembro de 2018, as competências que serão trabalhadas neste componente serão prioritariamente:

Competências técnicas: gerência, organização e recuperação das informações; proficiência em técnicas, habilidades e ferramentas emergentes; análise de dados para tomada de decisão.

Competências socioemocionais: trabalho em equipe; capacidade de lidar com conflitos; o aprendizado contínuo e colaborativo; e a liderança para a execução de ações nos diversos ambientes organizacionais.

9 Ementa

Princípios de Big Data. Hadoop e Armazenamento de Dados. Princípios de Desenvolvimento com PySpark. Análise de Dados em Python com Pandas. Big Data Analytics.

10 Objetivos

Criticar os processos de trabalho e obter interpretações rápidas e valiosas sobre tendências de mercado, baseando-se nos princípios de Big Data, para planejar o processamento de uma quantidade enorme de informações fazendo isso com alto desempenho e disponibilidade.

Esquematizar o armazenamento e processamento distribuído de grandes conjuntos de dados, utilizando a plataforma de código aberto Hadoop, para criar clusters de computadores com hardware commodity obtendo armazenamento, processamento, acesso, governança, segurança e operações de dados.

Construir aplicações de computação paralela, tendo como ferramenta o PySpark, para criar aplicações com pipeline de aprendizado de máquina realizando previsões nos dados.

Analisar grandes volumes de dados, utilizando a biblioteca Pandas, para propor a utilização de técnicas avançadas de análise de dados.

Esquematizar modelo de análises de dados estruturados e semiestruturados, baseando-se em técnicas de descoberta do conhecimento, para construir algoritmos de inteligência artificial.

11 Objetivos sociocomunitários

Implantar e configurar ambiente, baseado em ambiente em nuvem, para permitir a extração, tratamento, carga e análise de grandes massas de dados.

Analisar dados, com base em evidências extraídas de grandes massas de dados, para viabilizar tomadas de decisão mais assertivas das partes interessadas.

Apoiar e aprimorar o processo de tomada de decisão em organizações públicas e privadas, baseando-se em análise de dados, para trazer ganho de qualidade e de eficiência.

Habilitar as partes interessadas, com base nos procedimentos de operação da solução desenvolvida, para que desfrutem dos benefícios da tecnologia de Big-Data.

Prever fenômenos relevantes com antecedência suficiente, baseando-se em análise de dados, para viabilizar antecipação na tomada de decisão e realização.

12 Descrição do público envolvido

O público externo à IES e implicado na ação proposta, denominada parte interessada, é composto por: instituições públicas, privadas ou do terceiro setor, que possam se beneficiar da organização, registro e análise computacional de grandes conjuntos de dados para revelar padrões, tendências e associações, especialmente relacionadas ao comportamento e interações humanas.

13 Justificativa

De acordo com os artigos 3º e 6º do Capítulo I da Resolução CNE/CES nº 7, de 18 de dezembro de 2018, a Extensão na Educação Superior Brasileira ao integrar a matriz curricular e à organização de pesquisa, promove, em um processo interdisciplinar, a formação integral do aluno, através da aprendizagem por projetos, que estabelece um diálogo construtivo e transformador com diferentes setores da sociedade brasileira e internacional. Esse componente na formação do aluno justifica-se pela importância de promover a atuação da comunidade acadêmica e técnica, a partir das demandas sociocomunitárias onde se encontra a IES, para o enfrentamento das questões da sociedade brasileira, inclusive por meio do desenvolvimento econômico, social e cultural.

Do ponto de vista acadêmico, para que seja possível desenvolver este projeto, será necessária ampla articulação entre teoria e prática através de uma visão interdisciplinar, dialogando com disciplinas como Matemática e Lógica, Introdução à Segurança da Informação, Pensamento Computacional, Fundamentos de Redes de Computadores, Introdução à Programação de Computadores, Estrutura de Dados, Desenvolvimento Web em HTML5, CSS, Javascript e PHP, Paradigmas de Linguagem de Programação em Python, Programação Orientada a Objeto, Computação em Nuvem, Engenharia de Software e Segurança Cibernética.

O grupo de alunos precisará contextualizar a aplicação do conteúdo acadêmico envolvido no projeto à realidade da parte interessada. Para que isso seja possível, serão necessários habilidades socioemocionais como integridade, cordialidade, trabalho em equipe, mediação de conflitos, comunicação assertiva, escuta ativa, e liderança.

No Brasil, um país em que a tradição em tecnologia ainda está em construção, as organizações frequentemente não conhecem a realidade de Big Data, e seus potenciais benefícios, ou não possuem competências para desenvolver soluções computacionais para extrair valor destes dados.

A análise de big data ajuda as organizações a aproveitar seus dados e usá-los para identificar novas oportunidades. Isso, por sua vez, leva a movimentos de negócios mais inteligentes, operações mais eficientes, e maior satisfação dos clientes. As empresas que usam big data com análises avançadas ganham valor de várias maneiras.

As tecnologias de big data, como análises baseadas em nuvem, podem reduzir significativamente os custos quando se trata de armazenar grandes quantidades de dados. Além disso, a análise de big data ajuda as organizações a encontrar formas mais eficientes de fazer negócios, ou mesmo para implementar seus processos. Isso gera até mesmo impacto positivo sobre o meio ambiente.

Adicionalmente, promove-se nas organizações a capacidade de analisar as informações imediatamente após a sua geração, e tomar decisões rápidas e informadas, o que aumenta a sua disposição de se adaptar a mudanças no comportamento de seus clientes. Pode ser possível, também, avaliar as necessidades e a satisfação do cliente, o que oportuniza a entrega de produtos mais alinhados com as expectativas. Com a análise de big data, aumenta-se a viabilidade de se desenvolver produtos inovadores para atender às necessidades em constante mudança.

O projeto extensionista desta disciplina da disciplina visa atuar na transformação dessa realidade, através da articulação teoria-prática multidisciplinar contextualizada à realidade das partes interessadas e canalizada por princípios socioemocionais que propiciem uma dialógica amistosa e mutuamente benéfica.

Esta disciplina adota a metodologia de aprendizagem baseada em projetos, construídos de forma dialógica com a sociedade (representada pela parte interessada) de forma atender aos objetivos citados anteriormente.

O estudo das teorias e práticas previstas na disciplina para a aptidão de desenvolvimento deste projeto extensionista, serão realizadas em paralelo com as atividades extensionistas. Assim, é importante o docente esteja ciente de que a metodologia de sala de aula invertida é a indicada para que o aluno possa estar preparado para as discussões técnicas e sociais envolvidas no processo que serão realizadas nos encontros em sala de aula. Dessa forma viabiliza-se o desenvolvimento concomitante das habilidades técnicas e do projeto extensionistas previstos para esta disciplina.

Sugere-se que a turma seja dividida em grupos (com no máximo 5 alunos por grupo), de acordo com a realidade no semestre corrente. Cada grupo será responsável por definir seu parceiro externo no projeto e de realizar pesquisa e contato, sob orientação do docente. Importante que todas as etapas sejam devidamente.

Uma curadoria de materiais prévia é imprescindível para as reflexões e aprendizado do conteúdo da disciplina, podendo-se selecionar artigos de jornais, revistas ou até acadêmicos que forneçam embasamento das teorias e tecnologias expostas nesta disciplina e que exponham tópicos de big data em python.

Durante o desenvolvimento de seus projetos extensionistas os grupos devem ir gradativamente preenchendo o Roteiro de Extensão, conforme modelo disponibilizado. Este roteiro serve como guia dos passos a serem seguidos pelos alunos, sob orientação do docente, e inclui descrições sucintas sobre o que cada seção deve conter. Os prazos para as entregas de cada seção do Roteiro de Extensão devidamente preenchida (um por grupo) serão definidos pelo docente e refletidas no cronograma de plano de trabalho presente no Roteiro de Extensão. Isso é fundamental para o efetivo acompanhamento do desenvolvimento do projeto durante o semestre letivo.

Etapas do desenvolvimento do projeto

1. Definição do grupo de trabalho e parte interessada envolvida:

1.1. Em sala de aula, montar o grupo de trabalho.

1.2. Identificar as partes interessadas: fazer contato com a comunidade e verificar o interesse na participação do projeto. Nesse contato pode ser necessária uma ou mais visitas ao local pretendido

1.3. Realizar o diagnóstico das demandas necessárias e definir o escopo e a priorização

1.4. Definição de cronograma de atuação/ visitas às partes interessadas

2. Diagnóstico e teorização do projeto, conforme Seção 1 do modelo de Roteiro de Extensão.

3. Planejamento e desenvolvimento do projeto, conforme Seção 2 do modelo de Roteiro de Extensão

4. Detalhamento técnico do projeto, a ser incluído na Seção 2.6 - Detalhamento técnico do projeto do Roteiro de Extensão
 - 4.1. Documentar um projeto de Tópicos de Big Data em Python que esteja alinhado aos objetivos e objetivos socio-comunitários deste plano de aprendizagem
 - 4.1.1. Definir o escopo do Projeto
 - 4.1.2. Escrever os Requisitos Funcionais do sistema
 - 4.1.3. Modelar as classes do sistema
 - 4.1.4. Contato com a parte interessada para o processo de validação da proposta
 - 4.2. Desenvolver o sistema
 - 4.2.1. Desenvolver as classes descritas
 - 4.2.2. Criar as interfaces do Sistema
 - 4.3. Testar e implantar o sistema
 - 4.3.1. Realizar testes de unidade nas classes do sistema
 - 4.3.2. Realizar o teste de integração com todas as partes do sistema
 - 4.3.3. Implantar o sistema na instituição
 - 4.3.4. Treinar o usuário responsável
5. Encerramento do Projeto, conforme Seção 3 do modelo de Roteiro de Extensão, incluindo Relato coletivo do grupo, avaliação de reação pela parte interessada dos resultados alcançados e relato de experiência individual de cada aluno do grupo.
6. Apresentação dos Projetos: nesta etapa, cada grupo deve apresentar o projeto e os resultados obtidos em cada trabalho. Para garantir a participação de todos os alunos no grupo, o docente pode definir um plano de ação, combinado anteriormente com a turma. O docente deve avaliar a entrega coletiva, a entrega individual e a avaliação de reação da parte interessada, presentes no Roteiro de Extensão, quanto à eficiência e eficácia do atingimento dos objetivos do projeto.

14 Procedimentos de ensino-aprendizagem

Esta disciplina adota a metodologia de aprendizagem baseada em projetos, construídos de forma dialógica com a sociedade (representada pela parte interessada) de forma atender aos objetivos citados anteriormente.

O estudo das teorias e práticas previstas na disciplina para a aptidão de desenvolvimento deste projeto extensionista, serão realizadas em paralelo com as atividades extensionistas. Assim, é importante o docente esteja ciente de que a metodologia de sala de aula invertida é a indicada para que o aluno possa estar preparado para as discussões técnicas e sociais envolvidas no processo que serão realizadas nos encontros em sala de aula. Dessa forma viabiliza-se o desenvolvimento concomitante das habilidades técnicas e do projeto extensionistas previstos para esta disciplina.

Sugere-se que a turma seja dividida em grupos (com no máximo 5 alunos por grupo), de acordo com a realidade no semestre corrente. Cada grupo será responsável por definir seu parceiro externo no projeto e de realizar pesquisa e contato, sob orientação do docente. Importante que todas as etapas sejam devidamente.

Uma curadoria de materiais prévia é imprescindível para as reflexões e aprendizado do conteúdo da disciplina, podendo-se selecionar artigos de jornais, revistas ou até acadêmicos que forneçam embasamento das teorias e tecnologias expostas nesta disciplina e que exponham tópicos de big data em python.

Durante o desenvolvimento de seus projetos extensionistas os grupos devem ir gradativamente preenchendo o Roteiro de Extensão, conforme modelo disponibilizado. Este roteiro serve como guia dos passos a serem seguidos pelos alunos, sob orientação do docente, e inclui descrições sucintas sobre o que cada seção deve conter. Os prazos para as entregas de cada seção do Roteiro de Extensão

devidamente preenchida (um por grupo) serão definidos pelo docente e refletidas no cronograma de plano de trabalho presente no Roteiro de Extensão. Isso é fundamental para o efetivo acompanhamento do desenvolvimento do projeto durante o semestre letivo.

Etapas do desenvolvimento do projeto:

1. Definição do grupo de trabalho e parte interessada envolvida:

1.1. Em sala de aula, montar o grupo de trabalho.

1.2. Identificar as partes interessadas: fazer contato com a comunidade e verificar o interesse na participação do projeto. Nesse contato pode ser necessária uma ou mais visitas ao local pretendido

1.3. Realizar o diagnóstico das demandas necessárias e definir o escopo e a priorização

1.4. Definição de cronograma de atuação/ visitas às partes interessadas

2. Diagnóstico e teorização do projeto, conforme Seção 1 do modelo de Roteiro de Extensão.

3. Planejamento e desenvolvimento do projeto, conforme Seção 2 do modelo de Roteiro de Extensão

4. Detalhamento técnico do projeto, a ser incluído na Seção 2.6 - Detalhamento técnico do projeto do Roteiro de Extensão

4.1. Documentar um projeto de sistema de controle de clientes e produtos/serviços, que esteja alinhado aos objetivos e objetivos socio-comunitários deste plano de ensino

4.1.1. Definir o escopo do Projeto

4.1.2. Escrever os Requisitos Funcionais do sistema

4.1.3. Modelar as classes do sistema

4.1.4. Contato com a parte interessada para o processo de validação da proposta

4.2. Desenvolver o sistema de controle de clientes e produtos/serviços

4.2.1. Desenvolver as classes descritas

4.2.2. Criar as interfaces do Sistema

4.3. Testar e implantar o sistema

4.3.1. Realizar testes de unidade nas classes do sistema

4.3.2. Realizar o teste de integração com todas as partes do sistema

4.3.3. Implantar o sistema na instituição

4.3.4. Treinar o usuário responsável

5. Encerramento do Projeto, conforme Seção 3 do modelo de Roteiro de Extensão, incluindo Relato coletivo do grupo, avaliação de reação da parte interessada e relato de experiência individual de cada aluno do grupo.

6. Apresentação dos Projetos: nesta etapa, cada grupo deve apresentar o projeto e os resultados obtidos em cada trabalho. Para garantir a participação de todos os alunos no grupo, o docente pode definir um plano de ação, combinado anteriormente com a turma. O docente deve avaliar a entrega coletiva, a entrega individual e a avaliação de reação da parte interessada, presentes no Roteiro de Extensão, quanto à eficiência e eficácia do atingimento dos objetivos do projeto.

15 Temas de aprendizagem

1. PRINCÍPIOS DE BIG DATA

1.1 INTRODUÇÃO E APLICAÇÕES AO BIG DATA

1.2 CONCEITOS DE IOT E COMPUTAÇÃO DISTRIBUÍDA

1.3 PLATAFORMAS EM NUVEM PARA APLICAÇÕES DE BIGDATA

1.4 PROCESSAMENTO E STREAMING DE DADOS

2. HADOOP E ARMAZENAMENTO DE DADOS

- 2.1 INTRODUÇÃO E ARQUITETURA AO HADOOP
- 2.2 ECOSSISTEMA E SOLUÇÕES COM HADOOP
- 2.3 HDFS VS RDBMS
- 2.4 ENTENDO UM DATA LAKE

- 3. PRINCÍPIOS DE DESENVOLVIMENTO DE SPARK COM PYTHON
 - 3.1 INTRODUÇÃO AO SPARK
 - 3.2 UTILIZANDO PYSPARK
 - 3.3 OPERAÇÕES DE MAPREDUCE COM PYSPARK
 - 3.4 TRANSFORMAÇÕES COM PYSPARK

- 4. ANÁLISE DE DADOS EM PYTHON COM PANDAS
 - 4.1 COMPONENTES E SINTAXE DO PANDAS
 - 4.2 PREPARAÇÃO DE DADOS COM PANDAS
 - 4.3 MANIPULAÇÃO DE DADOS COM PANDAS
 - 4.4 VISUALIZAÇÃO DE DADOS COM PANDAS

- 5. BIG DATA ANALYTICS (ATIVIDADE PRÁTICA SUPERVISIONADA)
 - 5.1 COMPREENDENDO A DESCOBERTA DO CONHECIMENTO - KDD
 - 5.2 DESMISTIFICANDO INTELIGÊNCIA ARTIFICIAL
 - 5.3 APLICAÇÕES DE APRENDIZADO DE MÁQUINA COM TENSOR-FLOW
 - 5.4 APLICAÇÕES DE APRENDIZADO PROFUNDO COM O SCIKIT-LEARN

16 Procedimentos de avaliação

O processo de avaliação se dá através de NOTA FINAL ÚNICA (NF), estabelecida ao fim do semestre. Os procedimentos de avaliação contemplarão as competências desenvolvidas no componente curricular, bem como os resultados dos projetos extensionistas. As avaliações poderão ser realizadas por meio de diversas atividades, definidas de acordo com o perfil do componente de extensão trabalhado. A soma de todas as atividades que possam vir a compor o grau final da NF não poderá ultrapassar o máximo de 10 (dez) pontos.

Os alunos serão divididos em grupos de trabalho e cada grupo de trabalho deverá elaborar à medida do desenvolvimento do projeto o Roteiro de Extensão seguindo o modelo disponibilizado.

Os itens a serem pontuados são as seções do Roteiro de Extensão, conforme modelo disponibilizado, a saber:

I. Diagnóstico e Teorização: 2,0 pontos,

II. Planejamento e Desenvolvimento do Projeto: 4,5 pontos, e

III. Encerramento do Projeto:

III.a. Relato Coletivo: 2,5 pontos,

III.b. Relato de experiência individual: 1,0 ponto (computado individualmente)

Os prazos para as entregas parciais do Roteiro de Extensão com as respectivas seções preenchidas (um por grupo) para acompanhamento do projeto serão definidos pelo docente e refletidas no cronograma de plano de trabalho presente no documento do grupo de trabalho. Cada aluno deve escrever seu relato individual, conforme seção 3.2 do Roteiro de Extensão, ou seja, o documento do grupo incluirá o relato individual de cada aluno.

O docente tem a prerrogativa de aplicar descontos de pontuação ou invalidar entrega em caso de atrasos.

Critérios avaliativos:

1. Desenvolvimento de atividades de acordo com o Plano de Aprendizagem;
2. Entregas das etapas materializadas no Roteiro de Extensão;
3. A articulação do conhecimento técnico e aplicação prática;
4. Cumprimento dos prazos estabelecidos no cronograma
5. A relevância social da atividade proposta

Para aprovação, o aluno deverá obter grau maior ou igual a 6,0, e ter frequência maior ou igual 75%.

17 Bibliografia básica

FACELI, Katti. **Inteligência Artificial Uma abordagem de aprendizado de máquina [BV:MB]**. 2ª Ed.. Rio de Janeiro: LTC, 2021.

Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788521637509/>

FREZATTI, Fábio. **Aprendizagem Baseada em Problemas**. São Paulo: Grupo GEN, 2018.

GOLDSCHMIDT, Ronaldo. **Data Mining - Conceitos, técnicas, algoritmos, orientações e aplicações [BV:MB]**. Rio de Janeiro: Elsevier, 2015.

Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788595156395/>

PEREIRA, Mariana Araújo. **Framework de Big Data [BV:MB]**. Porto Alegre: SAGAH, 2019.

Disponível em: <https://integrada.minhabiblioteca.com.br/books/9786556900803/>

18 Bibliografia complementar

BANIN, Sérgio Luiz. **Python 3 - Conceitos e Aplicações - Uma Abordagem Didática [BV:MB]**. São Paulo: Érica, 2018.

Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788536530253/>

Batista, M.D. C. [BV:MB]. **Data Mining**. Porto Alegre: SAGAH, 2021.

Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9786556900292/>

BENDER, William N. **Aprendizagem baseada em projetos: educação diferenciada para o século XXI**. Porto Alegre: Penso, 2014.

MELLO, Cleyson de Moraes; ALMEIDA, José Rogério Moura de Neto; PETRILLO, Regina Pentagna. **Curricularização da Extensão Universitária**. 2ª. Rio de Janeiro: Processo, 2022.

Disponível em: <https://plataforma.bvirtual.com.br/Leitor/Publicacao/198121/pdf/0?code=haO8b9eyXWALpYNVRvgcTpaKKTWSEC5yk7VHx8YkVA7x4ZpohYv3u0gjs>

MORAIS, Izabelly. **Introdução a Big Data e Internet das Coisas (IOT) [BV:MB]**. Porto Alegre: SAGAH, 2018.

Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788595027640/>

MUELLER, John Paul. **Aprendizado Profundo para Leigos [BV:MB]**. Rio de Janeiro: Alta Books, 2020.

Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788550816982/>

PERKOVIC, Ljubomir. **Introdução à Computação Usando Python - Um Foco no Desenvolvimento**

de Aplicações [BV:MB]. 1ª Ed. Rio de Janeiro: LTC, 2016.

Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788521630937/>