



PROJETO PEDAGÓGICO DE CURSO
Pós-Graduação *Lato Sensu*
MBA em Data Science

Sumário

1. Nome do Curso e Área do Conhecimento.....	3
2. Características Técnicas do Curso.....	3
3. Público-Alvo	3
4. Critérios de Seleção	4
5. Justificativa do Curso	4
6. Objetivos do Curso	4
7. Competências e Habilidades do Curso	5
8. Metodologia de Ensino e Aprendizagem	5
9. Estágio Não Obrigatório.....	8
10. Matriz Curricular	10
11. Carga Horária.....	11
12. Conteúdo Programático.....	11
13. Infraestrutura Física e Pedagógica.....	20

1. Nome do Curso e Área do Conhecimento

Nome do Curso: MBA em Data Science

Área de Avaliação (CAPES): CIÊNCIA DA COMPUTAÇÃO

Grande Área (CAPES): CIÊNCIAS EXATAS E DA TERRA

Área do Conhecimento (CAPES): CIÊNCIA DA COMPUTAÇÃO

Classificação OCDE: Computação e Tecnologias da Informação e Comunicação (TIC)

2. Características Técnicas do Curso

Modalidade: Educação a Distância

Número máximo de vagas por Polo/Unidade: 1000 alunos

Período de Oferecimento: O curso possui entrada intermitente, respeitadas as datas de início e de fim cadastradas na oferta, bem como observado o período indicado para a sua integralização.

Limitações legais

Resolução CNE/CES Nº 1, de 06 de abril de 2018, que estabelece normas para o funcionamento de cursos de pós-graduação lato sensu.

O candidato deverá ser graduado com diploma devidamente registrado segundo as normas estabelecidas pelo MEC.

3. Público-alvo

Engenheiros e cientistas da computação, analistas de sistemas, engenheiros de demais áreas, matemáticos, físicos, economistas ou administradores e demais profissionais com conhecimentos em tecnologia.

4. Critérios de Seleção

O ingresso na pós-graduação será realizado por meio de inscrição no portal da Instituição e entrega dos documentos solicitados. Em seguida, analisados pela área competente.

5. Justificativa do Curso

A ascensão do uso de dados no mercado de trabalho atual tem exigido que os novos profissionais saibam trabalhar com dados quantitativos e conheçam as ferramentas da ciência de dados. Essa exigência tem permitido que as empresas desenvolvam soluções mais inovadoras baseadas em dados coletados da sociedade que resultem em produtos mais eficientes e corretamente direcionados às pessoas, sobretudo as que interagem e contribuem com as mais recentes aplicações tecnológicas. O grande desafio para a formação desse profissional é unir habilidades técnicas e de negócio, permitindo que os profissionais com essas qualificações mantenham o foco no segmento corporativo ao passo que evoluem a análise quantitativa das informações disponíveis. Desta forma, o profissional do MBA em Data Science é habilitado a transformar dados em insights acionáveis, visando à melhoria da tomada de decisões estratégicas das empresas por meio da análise de dados. Este profissional é capaz de elaborar soluções e construir produtos inovadores observando a realidade da sociedade e das empresas através das ferramentas fornecidas pela ciência de dados.

6. Objetivos do Curso

6.1. Objetivo Geral

- Oferecer formação teórica e prática na aplicação de técnicas estatísticas e computacionais para a resolução de problemas de negócio e da vida social.

6.2. Objetivos Específicos

- Desenvolver visão de negócio orientada a dados e resultados;
- Capacitar na identificação e proposição de soluções técnicas aos problemas das organizações e da sociedade;
- Fornecer uma visão estratégia dos dados nas organizações;
- Disseminar práticas de mercado na formulação de problemas e soluções envolvendo Inteligência Artificial;
- Oferecer formação específica na área de ciência e engenharia de dados de maneira multidisciplinar, difundindo e disseminando o conhecimento dos fundamentos e da atividade prática.

7. Competências e Habilidades do Egresso

Competências

Atuar no desenvolvimento de produtos baseados em dados, coletando, transformando, gerenciando e modelando estatisticamente soluções em empresas de diferentes setores, incluindo indústrias e órgãos públicos.

Habilidades

- Identificar problemas de negócio capazes de serem resolvidos pelo uso eficiente de dados e tecnologias;
- Desenvolver produtos de dados a partir da pesquisa e desenvolvimento;
- Compreender técnicas, metodologias e algoritmos de Machine Learning;
- Analisar e manipular dados estruturados e não estruturados;
- Desenvolver aplicações usando linguagem de programação e ferramentas associadas;
- Auxiliar no processo de tomada de decisões estratégicas nas empresas e órgãos públicos.

8. Metodologia de Ensino e Aprendizagem

O desenvolvimento das disciplinas do curso se dará no ambiente virtual, onde o aluno cumprirá 40 horas por disciplina.

Estas 40 horas são compostas pelo estudo dos diversos insumos pedagógicos disponibilizados, como:

- ✓ materiais de leitura;
- ✓ videoaulas;
- ✓ slides;
- ✓ podcasts;
- ✓ indicações de leituras extras como artigos e capítulos de livros.

Como parte do modelo acadêmico, também contabilizamos:

- ✓ a interação com os tutores para esclarecimentos de dúvidas pedagógicas (via Fórum);
- ✓ a realização da avaliação disciplinar pelo aluno.

Vale ressaltar que cada aluno é único e tem seu próprio tempo para aprendizagem. Alguns conseguem assimilar o conteúdo com mais agilidade,

outros precisam retomar as leituras e videoaulas para tornar sua aprendizagem efetiva.

O tempo gasto pelo aluno para o entendimento de cada disciplina depende também da sua familiaridade com o tema, ou seja, o conhecimento prévio que o aluno tem em relação a temática abordada.

No ambiente virtual, o aluno encontrará o conteúdo das disciplinas, organizados em temas/webaulas.

Para cada um deles, o aluno realizará um conjunto de atividades baseadas em leitura de textos de fundamentação teórica e acesso a recursos audiovisuais.

Um tutor apoiará as atividades realizadas no ambiente virtual, atendendo o aluno nas suas dúvidas por meio de ferramentas de comunicação.

O aluno, ao iniciar os seus estudos, terá um encontro presencial para acolhida/ambientação; esse encontro terá como objetivos:

- ✓ Integrar o aluno ao curso de Pós-Graduação.
- ✓ Dialogar e esclarecer as dúvidas sobre a proposta pedagógica do curso e as regras acadêmicas.
- ✓ Apresentar ao aluno o Ambiente Virtual de Aprendizagem (o primeiro acesso; o envio de documentos; os serviços de secretaria e financeiro; a disciplina Ambientação; a tutoria online; o boletim acadêmico; as disciplinas e seus conteúdos; a biblioteca virtual; entre outros).
- ✓ Proporcionar um momento de Network aos pós-graduandos.

Avaliação do Desempenho do Aluno

O aluno deverá realizar as atividades propostas no ambiente virtual. A realização das atividades irá compor sua frequência no curso, que será considerada para a sua aprovação.

A atividade avaliativa que o aluno realizará para compor a sua média é a Avaliação Virtual (AV); essa atividade é obrigatória e estará disponível no

Ambiente Virtual de Aprendizagem – AVA, conforme cronograma de seu curso.

Para a aprovação em cada uma das disciplinas, o aluno deverá obter frequência de, no mínimo, 75% (setenta e cinco por cento) e nota igual ou superior a 7,0 (sete).

As notas devem ser expressas no intervalo de 0 (zero) a 10 (dez).

O aluno que obtiver média inferior a 7,0 (sete) nas disciplinas terá direito ao Programa de Dependência e Recuperação – PDR, mediante a solicitação de requerimento e respeitando o período de jubramento do curso.

O PDR será realizado no ambiente virtual de aprendizagem, sendo que o aluno terá acesso ao conteúdo da disciplina e realizará uma Avaliação Virtual - AV, e a nota obtida substituirá a média do aluno.

Para a obtenção do **Certificado** de Pós-graduação *Lato Sensu* – especialização, o aluno deverá cumprir todas as condições seguintes:

- ✓ Frequência mínima de 75% (setenta e cinco por cento) em todas as disciplinas;
- ✓ Nota igual ou superior a 7,0 (sete) em todas as disciplinas.

Certificação

O Certificado de conclusão de curso de Especialização será acompanhado por histórico escolar, em cumprimento às exigências da Resolução CNE/CES nº1, de 06 de abril de 2018, da Câmara de Educação Superior do Conselho Nacional de Educação.

Composição do Corpo Docente

O corpo docente do curso é constituído por profissionais qualificados, com comprovado saber em sua área de atuação, conforme Resolução CNE/CES nº1, de 06 de abril de 2018, sendo integrado, no mínimo, por 30% (trinta por cento) de portadores de título de pós-graduação *stricto sensu*, isto é, portadores de títulos de Mestrado e Doutorado, obtidos em programas de

pós-graduação *stricto sensu* devidamente reconhecidos pelo poder público em território nacional, ou revalidados, conforme legislação vigente. Os demais docentes são certificados em nível de especialização, pós-graduação *lato sensu*, de reconhecida capacidade técnico-profissional.

9. Estágio Não Obrigatório

O estágio curricular não obrigatório tem como finalidade estimular o aluno a desenvolver atividades extracurriculares, para que possa inter-relacionar os conhecimentos teóricos e práticos adquiridos durante o curso e aplicá-los na solução de problemas reais da profissão, proporcionando o desenvolvimento da análise crítica e reflexiva para os problemas socioeconômicos do país, de acordo com a Resolução de Estágio curricular não obrigatório vigente na Instituição.

O estágio curricular não obrigatório do curso de Pós-Graduação *Lato Sensu* segue as diretrizes estabelecidas pela Lei Nº 11.788/2008.

Os principais objetivos da prática do estágio curricular não obrigatório são:

- I. proporcionar o exercício do aprendizado compromissado com a realidade socioeconômica-política do país;
- II. propiciar a realização de experiências de ensino e aprendizagem visando à educação profissional continuada, alicerçada no desenvolvimento de competências e habilidades e ao exercício do pensamento reflexivo e criativo; e
- II. incentivar o trabalho de pesquisa e investigação científica, visando ao desenvolvimento da ciência, da tecnologia e da cultura.

A carga horária é definida pela concedente de estágio, não podendo ultrapassar a carga horária máxima de 6 (seis) horas diárias e 30 (trinta) horas semanais, as quais podem ser realizadas em empresas públicas ou privadas, instituição de pesquisa, órgãos governamentais e não governamentais, e as próprias unidades da Universidade, desde que sigam às condições adequadas para que o estagiário possa aprofundar os seus conhecimentos teóricos e práticos adquiridos no curso.

Para o Curso de Pós-Graduação EAD, a prática do estágio curricular não obrigatório é permitida durante a vigência do curso, não podendo exceder em um mesmo campo de estágio o período de 2 (dois) anos.

Os estágios curriculares não obrigatórios devem estar apoiados em Termo de Compromisso e de comum acordo com a Instituição, devendo explicitar não somente os aspectos legais específicos, como também os aspectos educacionais e de compromisso com a realidade social.

O Planejamento do Estágio Curricular Não Obrigatório é de responsabilidade do coordenador de curso/professor orientador e também do Departamento de Estágios.

10. Matriz Curricular

DISCIPLINAS	CH PRÁTICA	CH TEÓRICA	CH TOTAL
Ambientação	0 h	0 h	0 h
Análise de dados	0 h	40 h	40 h
Técnicas estatísticas: teoria e prática (R programming)	0 h	40 h	40 h
Análise e modelagem preditiva	0 h	40 h	40 h
Análise de redes sociais	0 h	40 h	40 h
Data Discovery, Olap e visualização de dados	0 h	40 h	40 h
Governança de dados	0 h	40 h	40 h
Linguagens de programação para ciência de dados_(Python com Spark)	0 h	40 h	40 h
Análise de agrupamentos	0 h	40 h	40 h
Machine learning	0 h	40 h	40h

11. Carga Horária

A carga horária de 360h constitui o conteúdo ministrado em 9 (nove) disciplinas.

12. Conteúdo programático

Disciplina: Ambientação

Ementa: Ensino a distância: características desta modalidade de estudo. A tecnologia e o ensino à distância. Legislação do Ensino a Distância no Brasil. Aspectos relacionados ao perfil no Ensino a Distância.

Competências e Habilidades:

- Compreender a modalidade de ensino a distância.
- Identificar uma relação entre tecnologia e educação no contexto do EAD;
- Discutir elementos da legislação do ensino a distância no Brasil;
- Refletir sobre o perfil do aluno em ensino a distância.

Conteúdo Programático 1: Introdução ao Ensino a Distância.

Conteúdo Programático 2: Tecnologia e educação.

Conteúdo Programático 3: Legislação do Ensino a Distância no Brasil.

Conteúdo Programático 4: Perfil do aluno no Ensino a Distância.

Bibliografia básica:

BRASIL. Ministério da Educação. **Lei de Diretrizes e Bases da Educação Nacional - Lei n. 9.394/96**. Brasília, 1996. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9394.htm. Acesso em: 6 dez. 2022.

BRASIL. Ministério da Educação. **Resolução CNE/CP n. 7, de 18 de dezembro de 2018**. Brasília, 2018. Disponível em: https://normativasconselhos.mec.gov.br/normativa/view/CNE_RES_CNECESN72018.pdf. Acesso em: 6 dez. 2022.

GIL, H. A passagem da Web 1.0 para a Web 2.0 e... Web 3.0: potenciais consequências para uma 'humanização' em contexto educativo. 2014. **Boletim informativo Cybercentro Castelo Branco**. Castelo Branco, 2014. Disponível em: <https://repositorio.ipcb.pt/bitstream/10400.11/2404/1/A%20passagem%20da%20Web%20Henrique.pdf>. Acesso em: 6 dez. 2022.

MEYER, A. I. da S. (2022). Conceituando a Educação a Distância. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 8, n. 1, p. 590-601, [s. l.], 2022. Disponível em: <https://periodicorease.pro.br/rease/article/view/3835>. Acesso em: 6 dez. 2022.

VAZ, M. L. de L.; RIBEIRO, F.; COSTA, L. A. da. Os desafios da educação a distância on-line e a remotividade na nova engenharia educacional. **Brazilian Journal of Science**, v. 1, n. 4, p. 79-86. Goiás, 2022. Disponível em: <https://www.brazilianjournalofscience.com.br/revista/article/view/79>. Acesso em: 6 dez. 2022.

VIEIRA, D. M. L.; COSTA, L. A.; VAZ, M. L. L. Um novo olhar para a educação a distância. In: MACHADO, G. E.; COSTA, S. C.; SILVA, K. R. P. **Debates contemporâneos: perspectivas e reflexões atuais**. Santa Maria: Arco, 2021.

Bibliografia Complementar:

BAPTISTA, M. M. Internet: auxílio à educação. **BIBLOS - Revista do Instituto de Ciências Humanas e da Informação**, v. 16, p. 37- 44, [s. l.], 2004. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/23717>. Acesso em: 6 dez. 2022.

MOORE, M.; KEARSLEY, G. **Educação a distância: uma visão integrada**. São Paulo: Thomson Learning, 2007.

MORAN, J. M.; MASETTO, M. T.; BEHRENS, M. A. **Novas tecnologias e mediação pedagógica**. 21. ed. Campinas: Papirus, 2013.

PIVA, D. J.; PUPO, R.; GAMEZ, L. et al. **EAD na prática: planejamento, métodos e ambientes de educação on-line**. São Paulo: Elsevier, 2011.

Disciplina: Análise de dados

Ementa: Apresentação do conceito de desenho de pesquisa como ponto de partida para o ensino de métodos quantitativos. A disciplina passará por

noções básicas de análise de dados e estatística, como a definição de população e amostra e construção de hipóteses. Também será feita uma introdução às inferências bivariadas e à distinção entre associação e causalção.

Competências e Habilidades:

- Ser consciente em relação à importância de um desenho de pesquisa robusto, assim como em relação às análises realizadas por meio dos métodos quantitativos;
- Ser capaz de realizar pesquisas empíricas com testes de hipótese e inferência bivariada;
- Saber a diferença entre associação entre duas ou mais variáveis e a ideia de causalção.

Conteúdo Programático 1: Desenho de pesquisa: o que é um desenho de pesquisa? E por que ele é importante? Noções de análise de dados: população e amostra; distribuições comuns; variáveis dependente, independente e de controle; mensuração de variáveis; confiabilidade e validade; erros de mensuração aleatórios e sistemáticos; estrutura dos dados.

Conteúdo Programático 2: Estatística descritiva: medidas de tendência central (média, moda e mediana); valores máximo e mínimo; medidas de dispersão (alcance, variância e desvio padrão); Introdução aos testes de hipótese: hipótese nula e hipótese alternativa.

Conteúdo Programático 3: Testes de hipótese bivariados: teste-t (para amostras independentes e emparelhadas); correlação; por que inferência não significa causalidade? Metodologia: aulas expositivas, com apresentações teóricas e práticas. As aulas apresentarão os métodos de maneira intuitiva, de forma a tornar a estatística mais amistosa aos que não são familiarizados com os temas.

Conteúdo Programático 4: As teorias serão apresentadas de forma aberta às intervenções dos discentes. Em termos práticos, serão utilizados dados reais e/ou de simulação para análise em *software* livre.

Bibliografia Básica:

AGRESTI, Alan; FINLAY, Barbara. **Métodos Estatísticos para as Ciências Sociais**. 4. Ed. Porto Alegre: Penso, 2012.

DANCEY, Christine P.; REIDY, JOHN. **Estatística sem Matemática para Psicologia**. 5. Ed. Porto Alegre: Penso, 2013.

KELLSTEDT, Paul M.; WHITTEN, Guy D. **Fundamentos da Pesquisa em Ciência Política**. São Paulo: Blucher. 2015.

TRIOLA, Mario. **Introdução à Estatística**. 12. ed. São Paulo: Grupo GEN. 2017.

WOOLDRIDGE, Jeffrey. **Introdução à Econometria**. Uma abordagem moderna. 3. Ed. São Paulo: Cengage Learning, 2017.

Bibliografia Complementar:

BOX-STEFFENSMEIER, Janet M.; BRADY, Henry E.; COLLIER, DAVID. **The Oxford Handbook of Political Methodology**. Oxford, UK: Oxford University Press, 2010.

BRADY, Henry E.; COLLIER, David (Editors). **Rethinking Social Inquiry: Diverse tools, shared standards**. 2nd ed. Lanham, MD: Rowman & Littlefield. 2010.

HAIR, Joseph F. et al. **Análise Multivariada de Dados**. 6. Ed. Porto Alegre: Bookman, 2009.

KING, Gary; KEOHANE, Robert O.; VERBA, Sidney. **Designing Social Inquiry: Scientific inference in qualitative research**. Princeton, NJ: Princeton University Press, 1994.

LATTIN, James; CARROLL, Douglas J.; GREEN, Paul E. **Análise de Dados Multivariados**. São Paulo: Cengage Learning, 2010.

POLLOCK, Philip. **The Essentials of Political Analysis**. 5th ed. Washington, DC: CQ Press, 2015.

VAN EVERA, Stephen. **Guide to Methods for Students of Political Science**. Ithaca, NY: Cornell University Press, 1997.

WHEELAN, Charles. **Estatística**. O que é, para que serve, como funciona. Rio de Janeiro: Zahar, 2016.

Disciplina: Técnicas estatísticas: teoria e prática (R

programming)

Ementa: A disciplina mostra os principais conceitos relacionados à estatística com o uso prático da linguagem R *Programming*.

Competências e habilidades

- Utilizar estatística em banco de dados.
- Compreender linguagem R.
- Verificar a utilização de gráficos em análise de dados.

Conteúdo Programático 1: Estatística e Linguagem R: Apresentação, Histórico e principais conceitos.

Conteúdo Programático 2: Estatística: Média, Mediana e Moda. Desvio Padrão e conceitos de Amostra. Intervalos de confiança, hipóteses, Correlação, Casualidade e Regressões lineares.

Conteúdo Programático 3: R: Principais comandos e estrutura.

Conteúdo Programático 4: Análise de Dados com a linguagem R.

Conteúdo Programático 5: Elaborando gráficos estatísticos com R.

Conteúdo Programático 6: Junção de Bancos de dados e sumarização estatística usando R.

Conteúdo Programático 7: Modelos Preditivos com R.

Bibliografia básica:

MARTINS, G. A. DOMINGUES, O. **Estatística Geral e Aplicada**. São Paulo: Atlas, 2017.

LARSON, R. FARBER, B. **Estatística Aplicada**. São Paulo: Pearson, 2017.

OLIVEIRA, F. E. M. **Estatística Geral e Aplicada**. São Paulo: LTC, 2017.

TRIOLA, M. F. **Introdução à Estatística**. São Paulo, SP, Brasil: LTC, 2017.

WHEELAN, C. **Estatística: O que é, para que serve, como funciona**. 1ed. Rio de Janeiro: Zahar, 2016.

MATLOFF, N. **The Art of R Programming: A Tour of Statistical Software Design**. San Francisco: No Starch Press, 2011.

Disciplina: Análise e Modelagem Preditiva

Ementa: Engenharia de atributos. Técnicas de pré-processamento de dados. Seleção de atributos. Avaliação de modelos preditivos. Técnicas de reamostragem. Otimização e seleção de modelos e hiperparâmetros. Overfitting e underfitting. Problemas de classificação.

Competências e Habilidades:

- Descrever técnicas de análise e modelagem preditiva
- Discutir problemas preditivos e suas soluções
- Revisar estratégias de classificação de dados
- Aplicar conceitos de análise preditiva e classificação em Python

Conteúdo Programático 1: Introdução à engenharia de atributos. Técnicas de pré-processamento de dados. Técnicas para eliminação de atributos, integração de dados, tratamento de dados desbalanceados, limpeza e transformação de dados. Abordagens para seleção de atributos.

Conteúdo Programático 2: Avaliação de modelos preditivos. Métricas de classificação (taxa de erro e acurácia). Matriz de confusão. Medidas de desempenho (recall, precisão, especificidade, AUROC). Técnicas de reamostragem (validação cruzada e holdout).

Conteúdo Programático 3: Processo de Indução de modelos. Generalização e regularização. Otimização de hiperparâmetros (random search e grid search). Overfitting e underfitting.

Conteúdo Programático 4: Problemas de classificação utilizando conjuntos de dados públicos. Aplicações tratando conjuntos de dados desbalanceados, detecção de outliers e anomalias usando linguagem Python (scikit-learn).

Bibliografia Básica:

FACELI, Katti et al. **Inteligência Artificial:** Uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2011.

de CASTRO, Leandro Nunes; FERRARI, Daniel Gomes. **Introdução à mineração de dados:** conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. New York: Springer, 2013.

Bibliografia Complementar:

BERGSTRA, James; BENGIO, Yoshua. Random search for hyper-parameter optimization. **The Journal of Machine Learning Research**, v. 13, n. 1, p. 281-305, 2012.

PROVOST, Foster; FAWCETT, Tom. **Data Science para negócios**. Tradução de Marina Boscatto. Rio de Janeiro: Alta Books, 2016.

GRUS, Joel. **Data Science do zero**: Primeiras regras com o Python. Alta Books, 2019.

Disciplina: Análise de Redes Sociais

Ementa: Redes complexas e aplicações. Redes sociais. Redes de informação. Redes biológicas. Redes tecnológicas. Grafos e conceitos. Medidas de redes. Coleta e extração de dados de mídias sociais na Web. Aplicações em redes sociais. Mineração de textos. Mineração de redes sociais. Detecção de comunidades.

Competências e Habilidades:

- Estudar cenários e aplicações de redes complexas;
- Discutir problemas em redes e suas soluções;
- Aplicar técnicas para a descoberta e mineração de dados em redes sociais;
- Empregar ferramentas para a análise de redes sociais.

Conteúdo Programático 1: Fundamentos das redes complexas. Conceitos introdutórios. Modelos de redes. Redes de mundo pequeno. Redes livre de escala. Redes aleatórias. Topologia das redes do mundo real. Tipos de redes: redes sociais, redes biológicas, redes de informação e redes tecnológicas.

Conteúdo Programático 2: Introdução à teoria dos grafos. Representação de redes sociais. Medidas estatísticas no nível do nó. Medidas estatísticas no nível da conexão. Análise de ligações. PageRank. Detecção de comunidades.

Conteúdo Programático 3: Ferramentas para análise e visualização de redes. iGraph. Networkx. Gephi.

Conteúdo Programático 4: Aplicações em redes. Extração de dados de mídias sociais. Mineração de textos na Web. Mineração de redes sociais.

Bibliografia Básica (inserir 3 bibliografias – atuais):

ALBERT, Réka; BARABÁSI, Albert-László. Statistical mechanics of complex networks. **Reviews of modern physics**, v. 74, n. 1, p. 47, 2002.

BARABÁSI, Albert-László. **Network science**. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, v. 371, n. 1987, 2013.

NEWMAN, Mark. The structure and function of networks. **Computer Physics Communications**, v. 147, n. 1-2, p. 40-45, 2002.

Bibliografia Complementar (inserir 3 bibliografias atuais):

BARABÁSI, Albert-László. **Linked: The new science of networks**. 2003.

NEWMAN, Mark. **Networks**. Oxford university press, 2018.

SIMÕES-PEREIRA, J. M. S. **Grafos e Redes: teoria e algoritmos básicos**. Rio de Janeiro: Interciência, 2013.

ZINOVIEV, Dmitry. **Complex network analysis in Python: Recognize-construct-visualize-analyze-interpret**. Pragmatic Bookshelf, 2018.

Disciplina: Data discovery, Olap e visualização de dados

Ementa: Compreensão da organização de dados, bem como sua visualização e ferramentas de visualização. Abordagem da ferramenta Olap e das ferramentas R, Python e Qlick Sense. Estudo das ferramentas de visualização Chart.js, Leaflet, Datawrapper, Dygraphs, Highcharts, Google Charts, Polymaps e Weka.

Competências e habilidades

- Compreender a visualização de dados em Big Data.
- Utilizar ferramentas de visualização.
- Executar descoberta de dados dentro de visualização.

Conteúdo Programático 1: Os Principais métodos de Visualização de Dados.

Conteúdo Programático 2: A Organização Visual (Visualização de dados e *Big Data analytics*).

Conteúdo Programático 3: O processo de Design de *Dashboard Design*.

Conteúdo Programático 4: Visualização de Dados com R, Python e Qlick Sense.

Conteúdo Programático 5: Visualização de Dados utilizando ferramentas Olap.

Conteúdo Programático 6: Data Discovery.

Conteúdo Programático 7: Outras ferramentas para visualização de Dados (Chart.js, Leaflet, Datawrapper, Dygraphs, Highcharts, Google Charts, Polymaps e Weka).

Bibliografia básica:

BARI, A. **Working with Big Data: Scaling Data Discovery**. Seattle, Washington, EUA: Amazon Digital Services LLC, 2017.

THOMSEN, E. **OLAP SOLUTIONS**. New York, NY, EUA: John Wiley & Sons, Inc, 2002.

GRUS, J. **Data Science do Zero: Primeiras regras com o Python**. Rio de Janeiro: Alta Books, 2016.

WRENBEEK, R. **Data Warehouses and OLAP: Concepts, Architectures and Solutions**. Treviso: Idea Group, 2006.

MALIK, S. **Enterprise Dashboards: Design and Best Practices for IT**. New York, NY, EUA: John Wiley & Sons, Inc, 2005.

BLOKDYK, G. **Visual Data Discovery Standard Requirements**. Austrália: Emereo Pty limited, 2018.

Disciplina: Governança de dados

Ementa: Visão geral da gestão de dados, governança de dados, gestão da arquitetura de dados, desenvolvimento de dados, gestão de dados mestre e dados de referência.

Competências e Habilidades:

- Compreender o ciclo de vida do dado e o papel da governança
- Elaborar projetos usando princípios de governança de dados
- Esquematizar arquitetura de dados
- Aplicar princípios da gestão de dados

Conteúdo Programático 1: Introdução à governança de dados. Dado como ativo corporativo. Ciclo de vida do dado. Gestão de dados. O profissional de governança de dados.

Conteúdo Programático 2: Visão geral da gestão de dados. Missão, princípios e objetivos. Funções e atividades. Papéis e responsabilidades. Tecnologias.

Conteúdo Programático 3: Conceito de governança de dados. Conceitos e atividades. Detalhes de atividades da governança de dados. Desenvolvimento de dados. Estilos de modelagem de dados. Design de dados. Implementação de dados.

Conteúdo Programático 4: Arquitetura de dados. Gestão da arquitetura de dados. Conceitos e atividades. Princípios gerais. Gestão de dados mestre. Conceitos de dado mestre e dado de referência. Integração de dados. Gestão de mudanças.

Bibliografia Básica:

BRACKETT, Michael; EARLEY, Production Susan. **The DAMA Guide to The Data Management Body of Knowledge** (DAMA-DMBOK Guide), 2009.

BARBIERI, Carlos. **Governança de Dados:** Práticas, conceitos e novos caminhos. Alta Books, 2020.

RÊGO, Bergson Lopes. **Gestão e governança de dados:** promovendo dados como ativo de valor nas empresas. Brasport, 2013.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business Intelligence e Análise de Dados para Gestão do Negócio.** Porto Alegre: Bookman, 2019.

Bibliografia Complementar:

HASHEM, I. A. T.; YAQOOB, I.; ANUAR, N.B.; MOKHTAR, S.; GANI, A.; KHAN, S.U. **The rise of “big data” on cloud computing: review and open research issues**. Information Systems 47 (2015), 98 - 115.

LOSHIN, David. **Master data management**. Morgan Kaufmann, 2010.

BAKSHI, Kapil. Considerations for big data: Architecture and approach. In: **2012 IEEE aerospace conference**. IEEE, 2012. p. 1-7.

Disciplina: Linguagens de programação para ciência de dados (*Python com Spark*)

Ementa: Apresentação da linguagem Python e sua utilização na análise de dados de grandes bases de dados. Uso em conjunto com Machine Learning. Integração com ferramentas como Spark e Hadoop

Competências e habilidades

- Utilizar linguagem Python.
- Compreender o ambiente de programação em Python
- Entender conceitos de Machine learning .

Conteúdo Programático 1: Introdução: Porque *Python*, *Spark* e *Hadoop*? e preparação do ambiente em *Spark* e *Python*.

Conteúdo Programático 2: Manipulação de dados com *Python*.

Conteúdo Programático 3: Organização e visualização de dados.

Conteúdo Programático 4: Análise estatística dos dados.

Conteúdo Programático 5: *Machine Learning* em *Python*.

Conteúdo Programático 6: Processando *Big Data* com *Spark*.

Conteúdo Programático 7: *Real Time Analytics* com *Python* e *Spark*.

Bibliografia básica:

MATTHES, E. **Curso Intensivo de Python: Uma Introdução prática e baseada em projetos à programação**. São Paulo: Novatec Editora, 2017.

CHAMBERS, B.; ZAHARIA, M. ***Spark: The Definitive Guide: Big Data Processing Made Simple***. San Francisco: O'Reilly Media, 2018.

MENEZES, N. N. C. **Introdução à Programação com Python: Algoritmos e Lógica de Programação Para Iniciantes**. São Paulo: Novatec Editora, 2017.

BEAZLEY, D.; JONES, B. K. ***Python CookBook: Recipes for Mastering Python 3***. 3ed. San Francisco: O'Reilly Media, 2013.

SLETKIN, B. **Python Eficaz: 59 maneiras de programar melhor em Python**. São Paulo: Novatec Editora, 2017.

Disciplina: Análise de Agrupamentos.

Ementa: Análise de agrupamentos. Critérios e índices de agrupamento. Algoritmos para a formação de grupos. Avaliação de grupos. Técnicas para preparação de dados. Aplicações em linguagem de programação.

Competências e Habilidades:

- Estudar aplicações de análise de agrupamentos;
- Apontar critérios para validação de modelos descritivos;
- Aplicar métodos de agrupamento em dados estruturados;
- Empregar ferramentas para a formação de grupos.

Conteúdo Programático 1: Conceitos e definições. Motivações. Aplicações no mercado. Processo de análise de agrupamentos. Representação de dados. Normalização. Medidas de similaridade e dissimilaridade. Avaliação de agrupamentos. Índices externos, internos e relativos. Validação de estruturas hierárquicas e particionais.

Conteúdo Programático 2: Métodos hierárquicos. Abordagem aglomerativa. Abordagem divisiva. Ligação simples, média e completa. Dendogramas. Aplicações em Python.

Conteúdo Programático 3: Métodos particionais. Algoritmo k-means. Algoritmo k-medians. Algoritmo k-modas. Algoritmo k-medóides. Aplicações em Python.

Conteúdo Programático 4: Métodos baseados em densidade. Algoritmo DENCLUE. Algoritmo DBSCAN. Algoritmo OPTICS. Aplicações em Python.

Bibliografia Básica:

AGGARWAL, Charu C.; REDDY, Chandan K. **Data clustering: algorithms and applications**. Chapman&Hall/CRC Data mining and Knowledge Discovery series, Londra, 2014.

FACELI, Katti et al. **Inteligência Artificial**: uma abordagem de aprendizado de máquina. 2. ed. Rio de Janeiro: LTC, 2021.

DE CASTRO, Leandro Nunes; FERRARI, Daniel Gomes. **Introdução à mineração de dados**: conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.

Bibliografia Complementar:

GAN, Guojun; MA, Chaoqun; WU, Jianhong. **Data clustering**: theory, algorithms, and applications. Society for Industrial and Applied Mathematics, 2020.

JOHNSON, Richard Arnold et al. **Applied multivariate statistical analysis**. London, UK: Pearson, 2014.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados**: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. Elsevier Brasil, 2017.

Disciplina: *Machine Learning*

Ementa: Algoritmos de aprendizado iterativo a partir dos dados. Modelos preditivos. Redes neurais e conceitos de *deep learning*. Sistemas de recomendação.

Competências e habilidades

- Conceber modelos preditivos.
- Conhecer métodos de decisão e modelos.
- Aplicar inteligência artificial.

Conteúdo Programático 1: Introdução ao *Machine Learning*. Como funciona a aprendizagem de máquina.

Conteúdo Programático 2: Algoritmos de *Machine Learning* e modelos preditivos.

Conteúdo Programático 3: Regressão, K-Nearest Neighbours (KNN) e Naives Bayes.

Conteúdo Programático 4: *Decision Tree*, *Random Forest* e Método Ensemble.

Conteúdo Programático 5: *Clustering*, *Support Vector Machines* e processamento em linguagem natural.

Conteúdo Programático 6: Redes neurais artificiais e introdução ao *Deep Learning*.

Conteúdo Programático 7: Sistemas de recomendação.

Bibliografia básica:

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. **Inteligência artificial: uma abordagem de aprendizado de máquina**. São Paulo: LTC Editora, 2011.

MITCHEL, M. B. **Machine Learning**. New York: McGraw-Hill Science/Engineering/Math, 1997. 432 p.

RASCHKA, Sebastian. **Python Machine Learning**. Birmingham: Packt Publishing, 2015.

FIGURELLI, R. **Laboratórios de robôs: inteligência evolutiva para Machine Learning**. Porto Alegre: Trajecta, 2017.

HOWARD, Mark. **Machine Learning and Deep Learning: Essential User Guide to Learn and Understand Machine Learning and Deep Learning Effectively**. Seattle: Amazon, 2017.

13. Infraestrutura Física e Pedagógica

O aluno encontrará todo o conteúdo do curso e assistirá às aulas gravadas no ambiente virtual. Para assistir às aulas é fundamental que as especificações abaixo sejam obedecidas, possibilitando, assim, uma recepção de maior qualidade dos vídeos.

Hardware:

- Processador Intel Core 2 Duo ou superior.
- 2Gb de Memória RAM.
- Placa de vídeo com resolução 1024x768, qualidade de cor 32 bits e compatível com Microsoft DirectShow.
- Microsoft DirectX 9.0c ou posterior.

Software:

- Navegador: Firefox, Google Chrome, Internet Explorer (sempre atualizado).
- Sistema Operacional: Windows XP ou posterior.
- Adobe Flash Player (atualizado).
- *Plugin* de vídeos SilverLigth (atualizado)

Rede:

- Conexão com a Internet banda larga de no mínimo 2 MB.
- Em caso de acesso em ambientes corporativos além da velocidade, é necessário verificar as condições de segurança de rede de sua empresa e se certificar que o site não estará bloqueado.

Adicionalmente, é prevista a utilização da biblioteca virtual para consultas bibliográficas e pesquisa de assuntos referentes às disciplinas ministradas.

