



PROJETO PEDAGÓGICO DE CURSO
Pós-Graduação *Lato Sensu*
Análise de Dados

Sumário

1.	Nome do Curso e Área do Conhecimento.....	3
2.	Características Técnicas do Curso	3
3.	Público-alvo	4
4.	Critérios de Seleção.....	4
5.	Justificativa do Curso.....	4
6.	Objetivos do Curso	5
7.	Competências e Habilidades do Egresso.....	5
8.	Metodologia de Ensino e Aprendizagem	6
9.	Estágio Não Obrigatório	9
10.	Matriz Curricular	10
11.	Carga Horária.....	10
1.	Conteúdo programático	11
12.	Infraestrutura Física e Pedagógica	26

1. Nome do Curso e Área do Conhecimento

Nome do Curso: Análise de dados

Área de Avaliação (CAPES): CIÊNCIA DA COMPUTAÇÃO

Grande Área (CAPES): 10000003 CIÊNCIAS EXATAS E DA TERRA

Área do Conhecimento (CAPES): 10300007 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

Profissionais de negócio, tecnólogos em gestão de marketing, comercial ou financeira, administradores, economistas, engenheiros, analistas de sistemas e demais profissionais de 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

Diversos segmentos de mercado – como marketing, finanças, seguros, varejo, agronegócio, tecnologia, etc. – têm requisitado novas competências dos profissionais que envolvam o desenvolvimento de soluções inovadoras e centradas na experiência dos clientes. Entre essas novas competências, estão aquelas relacionadas à análise de dados e ao uso das novas tecnologias. Embora muitos cursos de graduação trabalhem com matemática, estatística e computação, as empresas não têm conseguido recrutar profissionais que conheçam essas três áreas e ainda compreendam os fundamentos do negócio com o qual a empresa atua. A formação dos profissionais deve ser complementada além do que já foi cursada em nível de graduação para abranger novos conhecimentos além da área de formação, especialmente se ela não tiver sido na área técnica. Dessa forma, o curso de Análise de Dados vem suprir a ausência de profissionais interdisciplinares que sejam capazes de trabalhar com as ferramentas estatísticas e as novas tecnologias de Big Data. O profissional certificado neste curso fica habilitado a compreender as necessidades dos negócios, desenhar soluções analíticas e construir produtos

de dados usando metodologias e técnicas modernas para captura, análise e modelagem de dados.

6. Objetivos do Curso

6.1. Objetivo Geral

- Capacitar profissionais no uso de técnicas estatísticas, ferramentas para análise de dados e recursos computacionais com foco na resolução de problemas de negócio.

6.2. Objetivos Específicos

- Fornecer uma visão estratégica dos dados nas organizações;
- Estudar técnicas e ferramentas de estatística e análise de dados;
- Aplicar conceitos de análise de dados na formulação e solução de problemas de negócio;
- Utilizar ferramentas computacionais para analisar grandes volumes de dados;
- Oferecer formação específica na área de ciência e análise de dados de maneira interdisciplinar.

7. Competências e Habilidades do Egresso

Competências

Participar e conduzir pesquisas e análises de dados, promovendo o bom uso de ferramentas e tecnologias estatísticas com foco na adoção, extração,

integração, gestão, modelagem e apresentação de dados e informações gerenciais.

Habilidades

- Identificar problemas de negócio capazes de serem resolvidos pelo uso eficiente de dados e tecnologias;
- Conduzir pesquisas e análise de dados utilizando conceitos e técnicas da estatística;
- Analisar e manipular dados estruturados e não estruturados;
- Promover o uso de tecnologias de dados nas organizações, buscando pela eficiência dos recursos e a maximização das experiências dos clientes;
- Gerenciar dados e aplicações baseadas em dados;
- Auxiliar no processo de tomada de decisões nas empresas e outras entidades.

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 comprometido 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
Banco de dados relacional e big data	0 h	40 h	40 h
BI: a inteligência de negócios	0 h	40 h	40 h
Linguagem SQL para <i>Data Analytics</i>	0 h	40 h	40 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
Data Discovery, Olap e visualização de dados	0 h	40 h	40 h
Integração e fluxo de dados (ETL)	0 h	40 h	40 h
Governança de dados	0 h	40 h	40 h

11. Carga Horária

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

1. Conteúdo programático

Disciplina: Ambientação

Ementa: Histórico da Educação a Distância. Legislação da Educação a Distância no Brasil. Potencialidades da Educação a Distância. Flexibilidade de Acesso. Tecnologias para apoio à pesquisa. Aprendizagem colaborativa. Características do aluno na EAD. Boa convivência virtual: netiquetas.

Competências e Habilidades:

- Identificar as características do aluno da modalidade de educação a distância;
- Utilizar ferramentas tecnológicas de apoio a pesquisa;
- Comunicar-se com clareza e assertividade oralmente e por escrito;
- Gerenciar tempo e atividades de estudo.

Conteúdo Programático 1: Histórico da Educação a Distância.

Conteúdo Programático 2: Legislação da Educação a Distância no Brasil.

Conteúdo Programático 3: Potencialidades da Educação a Distância.

Conteúdo Programático 4: Flexibilidade de Acesso.

Conteúdo Programático 5: Tecnologias para apoio à pesquisa.

Conteúdo Programático 6: Aprendizagem colaborativa.

Conteúdo Programático 7: Características do aluno na EAD.

Conteúdo Programático 8: Boa convivência virtual: netiquetas.

Bibliografia básica:

GOTTARDI, M. de L. A autonomia na aprendizagem em educação a distância: competência a ser desenvolvida pelo aluno. **Associação Brasileira de Educação a Distância**, São Paulo, v. 14, n. 8, p. 110-123, dez, 2015.

MORAN, J. M. MASETTO, M. T. BEHRENS, M. A. **Novas Tecnologias e Mediação Pedagógica**. 21. ed. Campinas, SP: Papirus, 2013.

PIVA, D. J. PUPO, R. GAMEZ, L. OLIVEIRA, S. **EAD na Prática: Planejamento, métodos e ambientes de educação online**. São Paulo: Elsevier, 2011.

Bibliografia Complementar:

ALVES, L. Educação a distância: conceitos e história no Brasil e no mundo. **Associação Brasileira de Educação A Distância**, São Paulo, v. 10, n. 7, p.85-92, out. 2011.

MAIA, Carmen; MATTAR, João. **ABC da EaD: a educação a distância hoje**. São Paulo: Prentice Hall, 2007.

MOORE, Michael G., **Educação a distância: sistemas de aprendizagem on-line**. 3. ed., São Paulo: Cengage Learning, 2013.

Disciplina: Banco de dados relacional e big data

Ementa: Estruturação do Banco de Dados Relacional, Modelos de Entidade e Relacionamento, Conceitos de transações, Blocos de Código; Princípios da Linguagem SQL e seus Comandos. Introdução aos Sistemas de Gestão de Banco de Dados, bem como ao Big Data, repositórios de dados, Data Lake, e armazenamento em nuvem.

Competências e Habilidades:

- Compreender a importância do armazenamento de dados de forma adequada no modelo relacional do banco de dados;
- Estudar possíveis aplicações do modelo relacional do banco de dados;
- Aprender os princípios básicos da linguagem SQL;
- Entender as razões que calçam a crescente utilização do Big Data.

Conteúdo Programático 1: Banco de Dados e Big Data: União Necessária. Modelo Entidade Relacionamento; Estrutura de Banco de Dados Relacional;

Modelo Relacional; Bloqueio e Simultaneidade de Banco de Dados; Banco de Dados Relacional do Futuro: relacional e autônomo; introdução ao Big Data; ao Repositório de Dados; ao Data Lake e ao armazenamento em Nuvem.

Conteúdo Programático 2: ACID e Benefícios do Banco de Dados Relacional. Princípios de Atomicidade; Consistência de Dados; Integridade e Durabilidade do banco de dados Relacional; Conceito de transações; Cardinalidade entre as transações (cardinalidades máximas e mínimas); Benefícios dos Bancos de Dados Relacionais; Aplicações e Futuro do Banco de Dados Relacional.

Conteúdo Programático 3: A SQL e os Sistemas de Gestão de Banco de Dados. Compreensão das características básicas da Linguagem Estruturada de Pesquisa (SQL); principais comandos embasados em DDL e DML; vantagens e desvantagens do SQL; introdução aos Sistemas de Gestão de Banco de Dados no meio empresarial e a relação com a SQL.

Conteúdo Programático 4: Big Data e bancos de dados não relacionais. Compreendendo o conceito e aplicabilidade da Big Data; 5 “V” do Big Data; ferramentas; Cenário atual; gerenciamento de dados com Hadoop e NoSQL.

Bibliografia:

COSTA, R. L. SQL: **Guia Prático**. 2. ed. Rio de Janeiro: Brasport, 2006.

DATE, C. J. **SQL e Teoria Relacional: Como escrever códigos SQL precisos**. Tradução: Rafael Zanolli. São Paulo: Novatec, 2015.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de Banco de Dados**. Tradução: Marília G. Pinheiro. 4. ed. São Paulo: Pearson, 2009.

FERRARI, F. C. **Crie Banco de Dados em MySQL**. São Paulo: Digerati Books, 2007.

GRAVES, Mark. **Projeto de Banco de Dados com XML**. São Paulo: Pearson, 2003.

HEUSER, C. A. **Projeto de Bancos de Dados**. Porto Alegre: Editora Bookman, 2009.

LAUDON, K.; LAUDON, J. **Sistemas de informações gerenciais**. Tradução: Célia Taniwaki. 11. ed. São Paulo: Pearson, 2014.

MACHADO, F. N. R. **Projeto e implementação de banco de dados**. 4. ed. São Paulo: Érica, 2020.

TAURION, C. **Big Data: velocidade, volume, variedade, veracidade, valor.** 1. Ed. Rio de Janeiro: Brasport, 2013.

Disciplina: BI: a inteligência de negócios

Ementa: O caminho da tecnologia para o crescimento exponencial. Aplicação nos negócios. Negócios inteligentes, negócios cognitivos, inteligência de mercado. Tomadas de decisão dentro das empresas. Amparo maior em análise de dados. Assertividade. Inteligência de negócios: principais elementos, como são feitos, analisados e gerenciados.

Competências e Habilidades:

- Ser capaz de aplicar *business intelligence* para resolução de problemas da área de negócios;
- Ser capaz de articular mineração de dados a partir de inteligência artificial;
- Dominar modelos preditivos e descritivos com intuito de embasar a tomada de decisão;
- Confeccionar apresentação de dados de forma didática e atrativa.

Conteúdo Programático 1: Introdução ao BI e Data *warehouse*.

Conteúdo Programático 2: Análise de negócios e visualização de dados.

Conteúdo Programático 3: Mineração de dados e Inteligência Artificial.

Conteúdo Programático 4: Business Performance Management (BPM).

Bibliografia Básica:

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

TURBAN, E. *et al.* **Business intelligence:** um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009.

MARÓSTICA, E. (org.) **Inteligência de mercado.** São Paulo: Cengage Learning, 2014.

Bibliografia Complementar:

CHEN, Hsinchun; CHIANG, Roger HL; STOREY, Veda C. Business intelligence and analytics: From big data to big impact. **MIS quarterly**, v. 36, n. 4, 2012.

LAURSEN, Gert HN; THORLUND, Jesper. **Business analytics for managers: Taking business intelligence beyond reporting**. John Wiley & Sons, 2016.

NEGASH, Solomon. Business intelligence. **The communications of the Association for Information Systems**, v. 13, n. 1, p. 54, 2004.

Disciplina: Linguagem SQL para Data Analytics.

Ementa: Estudo das linguagens de definição e manipulação de dados, com ênfase na linguagem SQL. Aplicações da análise de dados usando ferramentas para a análise de negócios.

Competências e Habilidades:

- Adquirir conhecimentos teóricos e práticos relacionados ao processo de planejamento, construção e manutenção de banco de dados;
- Compreender a arquitetura de um sistema de banco de dados e seu funcionamento para análise de dados;
- Criar e manipular bancos de dados usando linguagem SQL;
- Utilizar tecnologia de gerenciamento de banco de dados para construir e manipular dados.

Conteúdo Programático 1: Projeto de banco de dados. Modelos de bancos de dados. Modelo Entidade-Relacionamento. Normalização de banco de dados. Sistemas Gerenciadores de Banco de Dados (SGBD). Apresentação de um SGDB para uso prático (MySQL, SQL Server, Oracle ou Big Query). Tipos de dados.

Conteúdo Programático 2: Linguagem de definição de dados (DDL). Comandos CREATE, ALTER e DROP para bancos de dados e tabelas.

Relacionamentos. Chaves primária e estrangeira. Linguagem de manipulação de dados (DML). Inserção, atualização e remoção de valores em bancos de dados.

Conteúdo Programático 3: Consultas em bancos de dados. Operadores lógicos e relacionais. Comando SELECT. Cláusula WHERE. Comando NULL. Cláusula ORDER BY. Funções numéricas. Funções de datas. Funções de caracteres e textos. Funções de agregação.

Conteúdo Programático 4: Consultas avançadas. Cláusula GROUP BY. Funções de agregação. Cláusula HAVING. Subconsultas. Junções.

Bibliografia Básica:

HEUSER, Carlos Alberto. **Projeto de banco de dados**. 6. ed. Porto Alegre: Bookman, 2009.

CARDOSO, Gisele Cristina; CARDOSO, Virginia Mara. **Linguagem SQL**. São Paulo: Saraiva Educação SA, 2017.

DAMAS, Luís. **SQL: Structured Query Language**. Rio de Janeiro : LTC, 2014.

Bibliografia Complementar:

PRICE, Jason. **Oracle Database 11G SQL**. Porto Alegre: Bookman, 2009.

LEBLANC, Patrick. **Microsoft SQL Server 2012: Série Passo a Passo**. Porto Alegre: Bookman, 2013.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados: fundamentos e aplicações**. 7. ed. São Paulo: Pearson Education do Brasil, 2018.

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: 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: Integração e fluxo de dados (ETL)

Ementa: A disciplina de Integração de Fluxo de Dados (ETL) mostra aos alunos conceitos importantes, como o processo de extração, transforma e carga de dados (ETL) para ambientes de Data Warehouse por meio de uma ferramenta visual de integração o SSIS (Integration Services) garantindo o total controle automatizado no processo de carregamento com o mínimo de intervenção humana possível.

Competências e habilidades

- Compreender a tecnologia ETL;
- Analisar e compreender as topologias da tecnologia ETL;
- Efetuar a obtenção de dados dentro do ambiente.

Conteúdo Programático 1: Introdução: Conceitos de extração, Transformação, carga, arquitetura, componentes, fluxo e integração de dados.

Conteúdo Programático 2: Ferramenta de ETL (SSIS – Integration Services da Microsoft): Componentes, design, operador, topologia, segurança, agente e repositório.

Conteúdo Programático 3: Criação de Topologias, esquemas físicos, esquemas lógicos e contextos.

Conteúdo Programático 4: Conceitos de modelos SSIS e data stores: projetos, ferramentas e linguagens principais utilizadas.

Conteúdo Programático 5: Criação Básica de ETL's para os modelos Bidimensionais desenhados na construção do DW (Data Warehouse) no tópico anterior.

Conteúdo Programático 6: Criação de ETL's intermediários para os modelos Bidimensionais criados na Construção do DW (Data Warehouse) tópico anterior.

Conteúdo Programático 7: Criação Avançada de ETL's para os modelos Bidimensionais criados na Construção do DW (Data Warehouse) tópico anterior. Integração e Controle do Fluxo de Dados nos modelos elaborados.

Bibliografia básica:

KIMBALL, R.; CASERTA, J. **The Data Warehouse ETL TOOLKIT**. Indianópolis, Indiana: Willey (Digital), 2009.

NOVAIS, R. C. N. **The Data Warehouse Toolkit**. São Paulo, SP, agosto de 2012. Disponível em: <<http://www.fatecsp.br/dti/tcc/tcc00071.pdf>>. Acesso em 25 de setembro de 2018.

MITCHELL, T.; MASSON, M.; LEONARD, A.; MOSS, J.; UFFORD, M. **SQL Server Integration Services Design Patterns**. New York, NY, EUA: Appess, 2014.

Bibliografia complementar:

GONÇALVES, R. R. **Integração de Dados na Prática. Técnicas de ETL Para Business Intelligence com Microsoft Integration Services 2012**. São Paulo: Érica, 2012.

MAFRA, Felipe. **Business Intelligence. Com o Sql Server 2012 – Guia Prático**. São Paulo: Instituto Alpha, 2017.

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.

12. 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.

