



Projeto Pedagógico de Curso

Curso de Pós-Graduação em

Business intelligence, *big data* & *analytics* - ciência de dados

Sumário

1.	Nome do curso e área do conhecimento.....	2
2.	Características técnicas do curso	2
3.	Público-alvo.....	2
4.	Critérios de seleção.....	2
5.	Justificativa do curso.....	3
6.	Objetivos do curso	3
7.	Competências e habilidades do curso	4
8.	Metodologia de ensino e aprendizagem	4
	Composição do Corpo Docente.....	6
9.	Matriz curricular.....	6
10.	Carga horária.....	7
11.	Conteúdo programático.....	7
12.	Infraestrutura física e pedagógica	17

1. Nome do curso e área do conhecimento

Nome do Curso: *Business intelligence, big data e analytics* - ciência de dados

Grande Área (CAPES): Ciências Exatas e da Terra (10000003)

Área de Avaliação (CAPES): Ciência da Computação

Área do Conhecimento (CAPES): Ciência da Computação (10300007)

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: 100 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 nº1, 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 com formação superior em Ciência da Computação, Engenharia de Computação, Sistemas de Informação, e tecnólogos da área de Tecnologia da Informação e outros cursos correlatos que atuam ou desejam atuar em Inteligência de Negócios e TI, coordenadores de projetos, gerentes de TI, analistas de negócios, analistas de redes sociais e mídias e outros profissionais com interesse em tecnologias para análise de bases de dados para apoio a tomada de decisão.

4. Critérios de seleção

O ingresso na pós-graduação será realizado por meio de processo seletivo descrito em Edital

(inscrição, seleção e matrícula).

5. Justificativa do curso

Com a explosão no volume de dados e informação, decorrente da evolução das Tecnologias de Informação e Comunicação (TICs), é perceptível o crescimento do mercado de *Business Intelligence* (BI), *Big Data* e *Business Analytics* tanto no Brasil como no mundo. Portanto as previsões de mercado para utilização destas tecnologias são muito promissoras. Isso, por consequência, já direciona para a demanda efetiva por um profissional denominado Cientista de Dados, com competências e habilidades para compreender as estratégias e necessidades de negócios, bem como, gerenciar, projetar e desenvolver soluções de análise de grandes volumes de dados.

O curso visa formar profissionais que possam atuar no mercado de BI e no crescente mercado da Ciência de Dados. Através de disciplinas teóricas e práticas, os alunos são conduzidos a estudar e experimentar conceitos, praticar os principais tópicos do BI, do *Big Data* e *Analytics*, vivenciar a aplicação de técnicas para desenvolvimento de projetos relacionados à Ciência de Dados, BI e *Big Data* com objetivo direcionado à geração de valor para os negócios.

6. Objetivos do curso

3.1 Objetivos gerais

- Identificar e compreender as diversas perspectivas da área de *Business Intelligence*;
- Formar profissionais capazes de analisar o estado da arte de *Big Data* e *Business Analytics* com forte embasamento conceitual e prático;
- Capacitar os participantes a projetar, desenvolver, gerenciar projetos de BI e *BIG DATA*, identificando e discutindo conceitos emergentes relacionados a tratamento de dados e verificar seus impactos em projetos de BI e *BIG DATA*.

3.2 Objetivos específicos

- Mostrar os principais processos entre ferramentas para a montagem de um DW (*Data Warehouse*) e também os processos a serem utilizados para carregar este DW a partir dos bancos de dados organizacionais;
- Fornecer visão geral de ferramentas que possam apoiar no tratamento de grande quantidade de dados provenientes da organização.
- Mostrar o uso de *Machine Learning* de forma teórica e prática com o uso de exemplos diversos relacionando-os;
- Fornecer visão geral do ambiente de banco de dados, na forma de proporcionar formas para a análise de dados.

- Proporcionar conhecimentos e competências aos participantes em técnicas, ferramentas e tecnologias relacionadas a BI;
- Capacitar os participantes na análise dos problemas empresariais e a projetar, desenvolver e gerenciar projetos que demandam técnicas atuais para análise de grandes volumes de dados, de maneira a apoiar a empresa para que ela alavanque sua competitividade;

7. Competências e habilidades do curso

Competências:

O curso visa formar profissionais que possuam conhecimentos de bancos de dados em arquitetura de data warehouses, tanto relacionais como não relacionais, que possam fazer análises de grande volume de informação, desenvolvendo algoritmos e utilizando ferramentas que tragam um desempenho computacional otimizado para a crescente quantia de dados. Também se espera que o egresso tenha conhecimento sobre o impacto das ferramentas de inteligência artificial neste ambiente.

Habilidades:

- Identificar e modelar banco de dados;
- Compreender DW (*Data Warehouse*);
- Analisar Bancos de dados Relacionais (SQL);
- Analisar Bancos de Dados não correlacionais (NoSQL);
- Identificar e compreender fluxo de dados;
- Desenvolver programação para *Big Data*;
- Compreender *Big Data*;
- Compreender e utilizar inteligência artificial;
- Compreender e utilizar *machine learning*.

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.

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:

- ✓ Leitura de textos de fundamentação teórica.
- ✓ Acesso às videoaulas.
- ✓ Aprofundamento dos seus conhecimentos, acessando elementos extratextuais.
- ✓ Desafio Profissional, com resolução para autoestudo.
- ✓ Realização de questões de autoestudo, para verificação de seu desempenho.

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 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. Matriz curricular

DISCIPLINAS	CH PRÁTICA	CH TEÓRICA	CH TOTAL
Ambientação	0h	0h	0h
Modelagem e arquitetura do dw (<i>data warehouse</i>)	0h	40h	40h
Integração e fluxo de dados (ETL)	0h	40h	40h
<i>Data discovery</i> , <i>olap</i> e visualização de dados	0h	40h	40h
Banco de dados relacional e <i>big data</i>	0h	40h	40h

Bancos de dados não relacionais (NoSQL)	0h	40h	40h
Técnicas estatísticas: teoria e prática (<i>R programming</i>)	0h	40h	40h
Linguagens de programação para ciência de dados (<i>Python com Spark</i>)	0h	40h	40h
<i>Machine learning</i>	0h	40h	40h
Projeto em ciência de dados com soluções para processamento paralelo e distribuído de dados	0h	40h	40h

10. Carga horária

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

11. 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:

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. Mensal.

Disponível em:

<http://www.abed.org.br/revistacientifica/Revista_PDF_Doc/2011/Artigo_07.pdf>. Acesso em: 19 fev. 2018.

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. Mensal. Disponível em:

< http://seer.abed.net.br/edicoes/2015/08_A_AUTONOMIA_NA_APRENDIZAGEM.pdf>
Acesso em: 19 fev. 2018.

LITTO, F. M. FORMIGA, M. M. M. (org.) **Educação a distância: o estado da arte**. São Paulo: Pearson Education do Brasil, 2009.

MOORE, Michael; KEARSLEY, Greg. **Educação a Distância, uma visão integrada**. São Paulo: Thompson Learning, 2007.

MORAES, M. C. **O Paradigma educacional emergente**. Campinas – SP: Papirus, 1997.

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

MORAN, J. M. **O Uso das Novas Tecnologias da Informação e da Comunicação na EAD - uma leitura crítica dos**

meios. <http://portal.mec.gov.br/seed/arquivos/pdf/T6%20TextoMoran.pdf>. Acesso em: 16 de fev de 2018.

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

Disciplina: Modelagem e arquitetura do DW (*data warehouse*)

Ementa: A disciplina de Modelagem e Arquitetura do DW (*Data Warehouse*) traz conhecimentos de modelos e arquiteturas condizentes com DW. Mostra ferramentas construção e verificar dados dentro das estruturas.

Competências e habilidades

- Reconhecer e modelar a arquitetura DW;
- Reconhecer e utilizar ferramentas em DW;
- Efetuar a obtenção de dados dentro do ambiente.

Conteúdo Programático 1: Banco de Dados Transacionais X Bancos de Dados Analíticos.

Conteúdo Programático 2: Conceitos básicos sobre *Data Warehouse*.

Conteúdo Programático 3: *Data Mart*.

Conteúdo Programático 4: Modelagem de Dados para um *Data Warehouse*.

Conteúdo Programático 5: Esquema Estrela e Esquema Floco de Neve.

Conteúdo Programático 6: Ferramentas de Dados em um *Data Warehouse*.

Conteúdo Programático 7: Mineração de Dados em *Data Warehouse*.

Bibliografia básica:

INMON, W. H. **Como Construir o *Data Warehouse***. Rio de Janeiro: Campus, 1997.

KIMBALL, R. ***The Data Warehouse Toolkit: guia completo para modelagem dimensional***. Rio de Janeiro: Campus, 2002.

Machado, F. N. **Tecnologia e Projeto de *Data Warehouse*: uma visão multidimensional**. São Paulo: Érica, 2004.

KIMBALL, Ralph. ***Data Warehouse – Construindo o Data Warehouse para Web***. Rio de Janeiro: Campus, 2000.

Bibliografia complementar

RASLAN, D. A.; CALAZANS, Angélica Toffano Seidel. ***Data Warehouse: Conceitos e Aplicações***.

Brasília, DF, Abril de 2014. Disponível em:

<<https://www.publicacoesacademicas.uniceub.br/gti/article/viewFile/2612/2400>>. Acesso em 25 set. 2018.

ZIULKOSKI, L. C. C. **Coleta de Requisitos e Modelagem de Dados para *Data Warehouse*: Um Estudo de Caso utilizando Técnicas de Aquisição de Conhecimento**. Porto Alegre, RS, Março de 2003.

Disponível em: <<http://www.inf.ufrgs.br/bdi/wp-content/uploads/Monografia-LuisZiulkoski.pdf>>. Acesso em 25 set. 2018.

NOVAIS, R. C. N. **Modelagem Dimensional**. São Paulo, SP, Agosto de 2012. Disponível em:

<<http://www.fatecsp.br/dti/tcc/tcc00071.pdf>>. Acesso em 25 set. 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: 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: Banco de dados relacional e *big data*

Ementa: Modelagem de entidade e relacionamento; Normalização; Operações com CRUD; Comandos SQL; Conceitos de transações. Princípios de segurança. Introdução ao *Big Data*. Integridade de dados.

Competências e habilidades

- Modelar banco de dados relacionais.
- Compreender comandos SQL.
- Verificar a utilização de segurança em banco de dados.
- Compreender integridade de dados.

Conteúdo Programático 1: Principais conceitos relacionados a *Big Data*. Metadados: *Big Data* e projeto físico para uma aplicação BI.

Conteúdo Programático 2: Modelo Entidade Relacionamento, modelagem entidade relacionamento, diagrama entidade relacionamento. Cardinalidade.

Conteúdo Programático 3: Modelo Relacional Normalizado. - 1ª Forma Normal. - 2ª Forma Normal. - 3ª Forma Normal.

Conteúdo Programático 4: Operações DDL: Create, show, describe, alter, drop, truncate e rename.

Conteúdo Programático 5: Operações DML: select, insert, update, delete.

Conteúdo Programático 6: Propriedades ACID - Atomicidade - Consistência – Integridade -

Durabilidade - Conceitos de transação.

Conteúdo Programático 7: Operações DCL (Direitos de acesso) - concessão de direitos (Grant) - revogação de direitos (Revoke) - papéis / perfis (Role)

Bibliografia básica:

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de banco de dados**. 4ed. São Paulo: Pearson, 2009.

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

MACHADO, Felipe Nery Rodrigues; ABREU, Mauricio Pereira de. **Projeto de banco de dados: uma visão prática**. 15. ed. São Paulo: Érica, 2008. 298 p.

Disciplina: Bancos de dados não relacionais (NoSQL)

Ementa: Conceituação de NoSql, prática e criação de bancos de dados não relacionais utilizando bancos comerciais de mercado como o MongoDB e o Hadoop. Estudo de formas de modelar bancos de dados não correlacionais.

Competências e habilidades

- Modelar banco de dados não correlacionais.
- Compreender comandos NoSQL.
- Verificar a utilização de segurança em banco de dados.
- Compreender integridade de dados.

Conteúdo Programático 1: Banco de Dados não-convencional: Introdução aos principais modelos de dados NoSQL;

Conteúdo Programático 2: NOSQL: Modelo de Sintaxe e dados JSON e XML.

Conteúdo Programático 3: Estrutura do MongoDB e Hadoop e seu uso prático.

Conteúdo Programático 4: Banco de Dados NoSQL: Modelo Orientado a documentos;

Conteúdo Programático 5: Banco de Dados NoSQL: Modelo Orientado a chave/valor;

Conteúdo Programático 6: Banco de Dados NoSQL: Modelo Orientado a família de colunas;

Conteúdo Programático 7: Banco de Dados NoSQL: Modelo Orientado a grafos;

Bibliografia básica:

SADALAGE, P. J.; FOWLER, M. **NoSQL essencial: um guia conciso para o mundo emergente da persistência poliglota**. 1ed. São Paulo: Novatec, 2013.

SOUZA, M. B. **Desvendando o MongoDB. Do Mongo Shell ao Java Driver**. 1ed. São Paulo: Ciência Moderna, 2015.

SOUZA, M. B. **Introdução ao MONGODB**. São Paulo: Novatec, 2015.

Bibliografia complementar:

PANIZ, D. **NoSQL: Como Armazenar os dados de uma aplicação moderna**. São Paulo: Casa do Código, 2016.

TRICANO, A. **MongoDB: NoSQL na prática**. São Paulo: André Tricano, 2015.

PICHILIANI, C. A. **Introdução do MongoDB: O ponto de partida para quem quer trabalhar com o banco de dados NoSQL orientado a documentos**. São Paulo: Amazon, 2015.

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: 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: *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.

Disciplina: Projeto em ciência de dados com soluções para processamento paralelo e distribuído de dados

Ementa: A disciplina traz o conceito de processamento paralelo e distribuído principalmente de consulta. Apresentar as arquiteturas existentes que trabalham atualmente com Sistemas Distribuídos.

Competências e habilidades

- Compreender arquitetura de processamento paralelo.
- Conhecer processamento distribuído.
- Aplicar conceitos do curso.

Conteúdo Programático 1: Introdução a Soluções para Processamento Paralelo e Distribuído de Dados. Gerenciamento de Clusters e Grids.

Conteúdo Programático 2: Monitoramento e depuração de programas paralelos. Sistemas de armazenamento de dados distribuídos.

Conteúdo Programático 3: Sistemas *Peer-to-Peers* para processamento Paralelo e Distribuído. Sistemas de Bancos de Dados Distribuídos principais características.

Conteúdo Programático 4: Construindo um *Data Warehouse* a partir de indicadores. Construindo ETL's para uso com o *Data Warehouse* criado.

Conteúdo Programático 5: Definindo um projeto de *Big Data* e as bases e arquivos a considerar. Tratando os arquivos de dados com *Python*.

Conteúdo Programático 6: Criando e tratando processos utilizando o *R Programming*. Criando e definindo *Data Lakes* a partir das bases disponibilizadas para tratativas.

Conteúdo Programático 7: Unificando e integrando os dados. Concluindo o projeto e exibindo os *Dashboards* em ferramentas OLAP.

Bibliografia básica:

KIMBALL, R. **The *Data Warehouse* Toolkit: guia completo para modelagem dimensional**. Rio de Janeiro: Campus, 2002.

BARI, A. **Integração de Dados na Prática. *Working with Big Data: Scaling Data Discovery***. Seattle: Amazon Digital Services LLC, 2017.

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

MAYER-SCHONBERGER, V. C. K. ***BIG DATA: Como Extrair Volume, Variedade, Velocidade e Valor da Avalanche de Informação Cotidiana***. Rio de Janeiro: Elsevier Editora Ltda, 2013.

TANENBAUM, A. S. **Sistemas Distribuídos: Princípios e Paradigmas**. 2ed. São Paulo: PEARSON. 2015.

DANTAS, M. **Computação Distribuída de Alto Desempenho: Redes, Clusters e Grids Computacionais**. Seattle: Amazon, 2017.

COLOURIS, G.; DOLLIMORE, J.; KINDBERG, T. **Sistemas Distribuídos: Conceitos e Projeto**. 5ed. São Paulo: Bookman, 2013.

SADALAGE, P. J.; FOWLER, M. **NoSQL essencial: um guia conciso para o mundo emergente da persistência poliglota**. 1ed. São Paulo: Novatec, 2013.

MATTHES, E. **Curso intensivo de Python: uma introdução prática e baseada em projetos à programação**. São Paulo: Novatec Editora, 2017.

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.

COLOURIS, G.; DOLLIMORE, J.; KINDBERG, T. **Sistemas distribuídos: conceitos e projeto**. 5ed. São Paulo: Bookman, 2013.

BAEZA-YATES, R.; RIBEIRO N. B. **Recuperação de informação: conceitos e tecnologia das máquinas de busca**. 2ed. São Paulo: Bookman, 2013.

CAMARGO, A.; AMARAL, E.; BASTOS, R. **Bancos de dados distribuídos: uma solução aplicada: Implemente e gerencie um SGBDD!** São Paulo: Novas Edições Acadêmicas, 2017.

VALDURIEZ, P.; OZSU, M. T. **Princípios de sistemas de bancos de dados distribuídos**. Rio de Janeiro: Elsevier, 2001.

DANTAS, M. **Redes de comunicação e computadores – abordagem quantitativa**. São Paulo:

Amazon.

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.

