

Programa (360 hs)

O programa do curso BI MASTER é composto por 12 disciplinas que são organizadas por 6 tópicos principais, como se segue:

SISTEMAS DE APOIO À DECISÃO

BI E BIG DATA

MACHINE LEARNING

INTELIGÊNCIA ARTIFICIAL

ESTATÍSTICA

MBA THESIS

A seguir detalhamos o programa de cada uma das 12 disciplinas do curso.

### **1. SISTEMAS DE APOIO À DECISÃO [SAD]**

Introdução a lógica de programação;

Introdução à Bancos de Dados Estruturados;

Dados Estruturados vs. Dados Não-Estruturados;

Modelo Relacional de Dados;

Linguagem SQL

Linguagem de Definição de dados (DDL);

Linguagem de Manipulação de dados (DML);

Comandos de Junção de tabelas (JOIN) e suas formas;

Elaboração de Consultas (DQL);

Funções de agregação de dados;

Subconsultas;

Elaboração de Visões

Diagramas Entidade-Relacionamento;

Programação em Python;

Introdução à linguagem Python;

Realização de operações matemáticas em Python (adição, multiplicação, subtração, divisão);

Apresentação de bibliotecas mais usadas (math, numpy, matplotlib, panda, system,);

Criação de funções pelo usuário;

Construção de loops com for e while;

Estruturas condicionais;

Uso de listas, vetores e dicionários;

Análise de Dados

Leitura de dados em Excel e arquivo.csv;

Geração de gráficos para análise de dados (seaborn e matplotlib);

Criação e manipulação de banco de dados em python (sql e sqlalchemy);

Introdução à biblioteca de processamento de linguagem natural (nltk);

## **2. BI e BIG DATA- Business Intelligence [BI]**

Introdução à Business Intelligence;

Transformação do conhecimento;

O Processo do BI;

Técnicas de Levantamento de Requisitos;

Tecnologia e Projeto de Data Warehouse (DW);

Introdução ao Data Warehouse;

Características de um DW;

Processos e Ferramentas de um DW;

Criação de um DW com Banco de Dados PostgreSQL;

Definição da Arquitetura do DW;

Tipos de Implementação de um DW;

Introdução a Modelagem Multidimensional;

Modelagem Transacional x Multidimensional;

Granularidade de Dados;

Tabela Fato;

Tabela Dimensão;

Hierarquias de Dimensão;

Medidas;

Modelo Estrela (Star Schema);

Modelo Floco de Neve (Snowflake);

Pontos Cardeais do BI;

Construindo Modelos com o Power Architect;

Introdução ao Projeto de ETL;

Extração de Dados;

Transformação de Dados;

Carga de Dados;

Ferramentas de ETL;

Chave Substituta;

Desnormalização;

Transformação e Carga de Dados Utilizando SQL;

Carga Incremental;

Agregação de Dados;

Pentaho Data Integration (PDI);

Carga de dados com PDI;

Transformação de dados com PDI;

Técnicas de Análise e Visualização de Dados;

Qualidade dos Dados;

Consultas Ad-Hoc X Mineração de Dados;

Ferramentas para Visualização de Dados;

Elaboração de Dashboards;

Design de Dashboards;

Introdução ao Power BI;

Fontes de Dados do Power BI;

Plotagem de Mapas;  
Outros Elementos do Power BI;  
Web Scrapping;  
Técnica Pivot;  
Categorização de Dados;  
Power BI e Banco de Dados;  
Conexão Via ODBC;  
Tratamento de Dados no Power Query;  
Linguagem DAX;  
Atualização de dados no Power BI;  
Criação de Medidas com linguagem DAX;  
Criação e Utilização de Indicadores;

Introdução a computação paralela;

Definição e conceitos;

FLOPS;

GPU vs. CPU;

Cluster e Grids;

Introducao a MapReduce;

Computação e processamento na nuvem: aplicações Windows Azure Amazon e Google;

Processamento paralelo em cluster na nuvem;

Text Mining;

Mídia Texto;

Tokenização;

Introdução para Processamento de Linguagem Natural;

Introdução a Web Mining;

Programação de de linguagem natural

Modelos embedding

Representações Bag of word TF-IDF

Representações de contexto: word2vec, Glove

Latent Dirichlet Allocation - LDA

Projeto de programação de linguagem natural

### **3. BI e BIG DATA - Localização e Uso da Informação [LUI]**

Big Data e Data Science;

Introdução a Big Data; Bases de dados não relacionais;

Hadoop Architecture;

MapReduce: Análise de Redes Sociais e Text Mining;

Web mining, Coleta de informações

Web crawlers (robôs virtuais),

API (Application Programming Interface);

Análise de sentimentos com informações de facebook, twitter, linkedin e outros sites;

Data Lakes;

Métodos de apoio à decisão em bases de dados;

Junção de SQL com Data Mining;

Aplicações de Datawarehouse;

Datawarehouse para Hive;

NoSQL;

NewSQL;

Apache Spark;

Definição e Arquitetura;

Bases de dados como Objetos;

Implementação e casos de uso;

RDD — Resilient Distributed Dataset;

Python para Big Data, Dados em Python;

Integração Spark Python;

Introdução ao PySpark;

#### **4. BI e BIG DATA - Confiabilidade [CONF]**

Cyber Security;

Vulnerabilidades e ataques: Backdoor, Ataques de negação de serviço e de acesso direto, Multivetor, ataques polimórficos, etc;

Confiança de dados

Tratamento de anomalias

Anomalias em bases de dados.

Estudo de casos usando Autoencoders.

Criptografia de modelos de aprendizado de máquina

Fundamentos de Hacking

#### **5. MACHINE LEARNING - Data Mining [DM]**

Introdução a DM e Análise exploratória de Dados;

Problemas Típicos de DM;

Etapas de um projeto de DM;

Análise exploratória na prática;

Pré-processamento de dados;

Tratamento de missing values;

Normalização;

Redução de dimensionalidade: seleção de atributos (Filtros, Wrappers, Agregação);

Balanceamento;

Outliers;

Métricas de avaliação

Aprendizado supervisionado;

Aplicações;

Apriori;

Support Vector Machine (SVM);

Árvores de Decisão;

Comitês;

Random Forest;

Regressão logística;

K nearest neighbors;

Associação;

FPGrowth;

Eclat;

Agrupamento;

Particionamento (Kmeans, Kmedoids);

Hierárquico (DIANA, AGNES);

Densidade (DBSCAN);

Regressão;

Regressão linear simples;

Regressão linear múltipla;

Regressão não linear simples;

Regressão não linear múltipla;

Séries Temporais com Métodos tradicionais ;

Naive Bayes;

Média Móvel;

Amortecimento exponencial;

Autoregressivo integrados de média móvel;

Auto regressivo não linear;

## **6. INTELIGÊNCIA ARTIFICIAL - Previsão de Séries e Inferência por Redes Neurais / Deep Learning [RN]**

Redes Neurais Artificiais:

Introdução,

Arquitetura e Algoritmos de Aprendizado;

Características básicas;

História;

Redes Perceptron;

Redes Multilayer Perceptron;

Funções de Ativação;

Tarefas de Aprendizado;

Algoritmos de Aprendizado;

Aprendizado de Hebb, Competitivo, Boltzmann, Backpropagation;

Classificação e Regressão;

Aprendizado Supervisionado;

Convergência e Generalização;

Problemas comuns: definição do tamanho da rede, paralisia da rede e mínimos locais;

Taxa de aprendizado e momento;

Redes Auto-organizáveis;

Introdução ao Deep Learning;

Características básicas;

Múltiplas camadas;

Aplicações;

Redes básicas;

Frameworks utilizados: Tensorflow;

Álgebra lineal básica;



Tensores;

Algoritmos de otimização

Gradiente descendente estocástico, Adam, Adagrad;

Comparação de desempenho dos algoritmos de otimização;

Redes Convolucionais;

Introdução a operações de Convolução;

Mídia imagem;

Hiperparâmetros da rede;

Função de ativação;

Configurações convolucionais conhecidas;

Modelos conhecidos : Alexnet, Googlenet;

Transfer Learning;

Definição e Arquitetura;

Ajuste fino;

Implementação e casos de uso;

Redes Recorrentes;

Definição e arquitetura;

LSTM;

Séries temporais;

Aplicações em programação de linguagem natural;

LSTM bidirecional, Generative Adversarial Networks;

Modelos de Segmentação Semântica

Modelos de Detecção de Objetos

Processamento de Linguagem Natural usando Deep Learning

Word Representation;

Visualização de dados multidimensionais usando Tensorflow;

Reconhecimento de relações entre palavras;

Reconhecimento de entidade;

Mecanismos de *Attention*

Modelos Transformer

Modelos Bidirectional Encoder Representations from Transformers (BERT)

Generative Pre-trained Transformer : GPT

## **7. INTELIGÊNCIA ARTIFICIAL - Sistemas Baseados no Conhecimento por Lógica Fuzzy [LF]**

Introdução a Lógica Fuzzy e conjuntos fuzzy;

Relações e composições de relações;

Aritmética fuzzy, intervalos fuzzy, modelos de aplicação e teoria dos conjuntos fuzzy;

Aprendizado Automático Fuzzy;

## **8. INTELIGÊNCIA ARTIFICIAL – Otimização por Algoritmos Genéticos [OAG]**

Inteligência Artificial em problemas de otimização;

Algoritmos Genéticos na otimização de problemas numéricos;

Algoritmos Genéticos e suas aplicações;

Representação, decodificação e avaliação de soluções;

Reprodução genética: seleção, cruzamento e mutação;

Otimização de problemas com restrições;

Tratamento de Restrições;

Otimização com múltiplos objetivos;

Avaliação e otimização multiobjetivo;

Ferramentas para aplicação de Algoritmos Genéticos;

## **9. INTELIGÊNCIA ARTIFICIAL – Otimização de Planejamento [OP]**

Problemas de planejamento e otimização combinatória;

Tratamento de restrições de precedência;

Otimização de problemas planejamento e programação;

Otimização de Problemas numéricos, de alocação, de ordem e logísticos;

## **10. ESTATÍSTICA - Métodos Estatísticos de Apoio à Decisão [MEAD]**

Distribuição de Probabilidade;

Variável aleatória;

Função distribuição de probabilidade;

Principais distribuições de probabilidade;

Valor Esperado;

Média;

Variância;

Covariância;

Correlação;

Viés;

Erro padrão;

Estatística para Análise de Dados;

Amostragem aleatória simples;

Amostragem aleatória estratificada;

Amostragem aleatória de agrupamento;

Amostragem aleatória sistemática;

Bootstrapping;

Regressão linear simples;

Regressão linear múltipla;

Regressão logística;

Estatística para Seleção de Modelos;

Teste Z;

Teste qui-quadrado;

MAPE;

RMSE;

Coeficiente de determinação, coeficiente de determinação ajustado;

Métodos forward, backward, stepwise;

Multicolinearidade (VIF);

### **11. ESTATÍSTICA - Apoio à Decisão sob Incerteza [ADI]**

Técnicas estatísticas de amostragem;

Tomada de decisão sob incerteza;

Blockchain e suas aplicações;

### **12. Projeto de Sistemas Inteligentes de Apoio à Decisão [PROJ]**

Trabalho de Conclusão de Curso;

Conceitos de Decisão;

Inovação: Caracterização e Desafios;

PoC: Proof of Concept: etapas e desenvolvimento;

Identificação do Problema e Diagnóstico da situação atual;

Definição dos objetivos;

Avaliação e análise dos dados disponíveis;

Definição de Escopo do Projeto;

Modelagem e Desenvolvimento;

Testes e Avaliação de Resultados;

O trabalho de conclusão aborda tema de interesse do aluno ou da empresa na qual trabalha, sob orientação dos professores do programa. O BI MASTER da PUC-Rio adota uma nova forma

de realização e apresentação do TCC, compatível com o atual mercado profissional tecnológico, o qual valoriza a experiência prática do profissional na inovação e no desenvolvimento de provas de conceito (PoC: Proof of Concept) que revelam a experiência real do profissional.

Na proposta do BI MASTER, o aluno identifica um tema/problema, propõe uma solução e, sob a orientação dos professores do curso, desenvolve uma prova de conceito que, junto com uma descrição textual da solução, fica disponível para demonstração na web através de plataformas como o GitHub.