



CURSO DE
PÓS-GRADUAÇÃO

ANÁLISE DE BIG DATA

- DATA ENGINEERING

FIA BUSINESS SCHOOL

Há mais de 40 anos, a FIA é reconhecida como uma das melhores escolas de negócios em educação executiva e consultoria, com soluções no Brasil e em vários outros países.

Criada em 1980, por iniciativa de professores do Departamento de Administração da FEA-USP, a FIA é uma entidade privada, sem fins lucrativos. Mantém atualmente um Acordo de Cooperação Técnica com a USP, mediante o qual executa projetos em parceria.

Essa simbiose traz benefícios e complementaridades para ambas as instituições, constituindo-se no diferencial mais importante da FIA em relação ao mercado. De um lado, dissemina-se conhecimentos em Administração com a realização de atividades de ensino, consultoria e pesquisa aplicada, de outro, alimenta-se as atividades acadêmicas de pesquisa e ensino da Universidade.



AcREDITADA
pela Amba



Filiada à
AACSB



Filiada à
EFMD



AcREDITADA
com nota máxima
pelo MEC

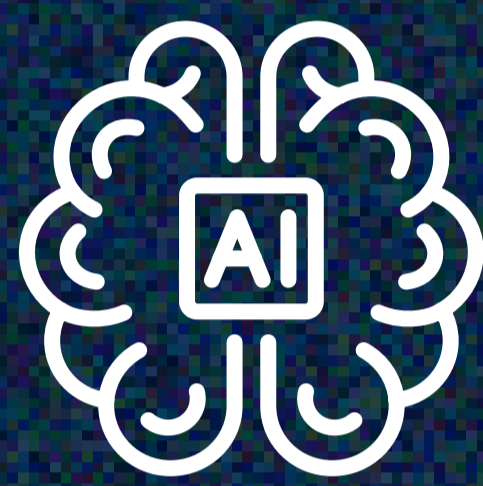


4 estrelas da
QS Stars 2022

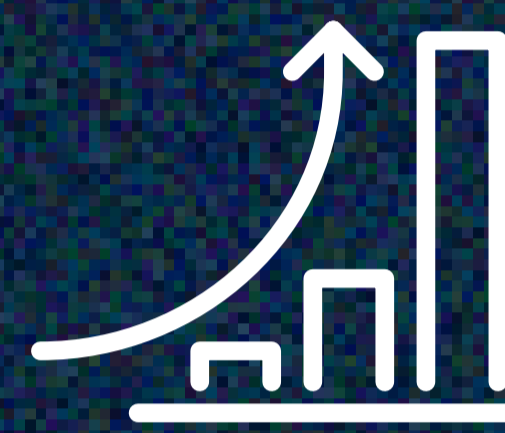
LABDATA

O **LABDATA – Laboratório de Análise de Dados** é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de *Big Data*, *Analytics*, Inteligência Artificial e Metaverso. É dirigido pela Prof^ª Dra. Alessandra de Ávila Montini.

Veja porque escolher o Labdata:



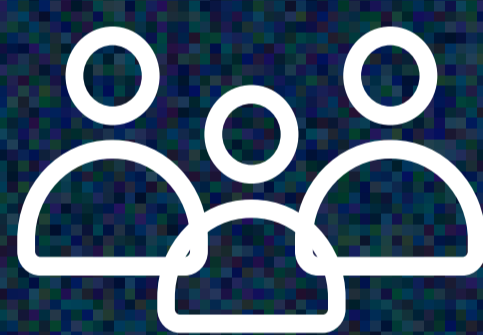
Centro de Excelência e pioneiro no lançamento dos cursos de *Big Data*, *Analytics* e Inteligência Artificial no Brasil



+ 10 anos de atuação
+ 15.000 alunos formados
+130 turmas de Pós e MBA



Lifelong Learning, atividades gratuitas como: *Hackathons*, Palestras, Cursos de Extensão e Eventos



Número limitado de alunos por turma.
Networking profissional qualificado e **painel de vagas** com as principais empresas



Cursos de Pós e MBA com carga horária mais completa do mercado, **plantão de dúvidas semanal** com o time de professores do Labdata



Conteúdo programático completo com aulas práticas e resolução de cases reais



Docentes com sólida formação acadêmica: doutores e mestres em sua maioria com larga experiência de mercado



Certificados reconhecidos pelo MEC e mesma certificação para os cursos online e presencial

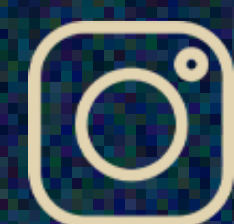


Localização de fácil acesso de carro ou transporte público

Redes Sociais



Labdata FIA



@labdatafia



Labdata - FIA

SOBRE O CURSO

OBJETIVO

- Apresentar de forma clara, objetiva e com aplicações reais como as metodologias de *Big Data* e Inteligência Artificial podem ajudar as empresas a obter vantagem competitiva;
- Apresentar como as tecnologias de *Big Data* e Inteligência Artificial podem ser utilizadas para a tomada de decisão e para a resolução de problemas complexos;
- Apresentar os principais componentes de *Big Data* utilizados em um projeto de engenharia de dados;
- Aplicar as tecnologias para tomada de decisão por meio de aulas teóricas, exercícios práticos, estudos de casos e projeto de *Big Data*.

PERFIL DO ALUNO

Profissionais que manipulam e precisam tomar rápidas decisões por meio de bases de dados e que tenham interesse em extrair informação por meio de modelos baseados em algoritmos de Inteligência Artificial e tecnologias de *Big Data*.

CORPO DOCENTE

O corpo docente conta com professores altamente capacitados com experiência no mundo corporativo. Nos critérios de seleção do corpo docente, serão priorizadas sua qualificação e experiências profissionais nas distintas matérias, de maneira que o curso permita não somente a transmissão de conhecimentos, mas também experiências enriquecedoras para os alunos.

METODOLOGIA

- As aulas na modalidade presenciais serão realizadas em um ambiente propício para aprendizagem;
- As aulas na modalidade EaD ao vivo serão transmitidas por meio de uma plataforma digital;
- Todas as aulas terão a presença de um professor titular com apoio de pelo menos um professor assistente;
- Serão utilizados diversos recursos tecnológicos para o melhor entendimento dos conceitos;
- Todas as aulas serão gravadas e permanecerão disponíveis em nossa biblioteca digital até o final do curso.

REQUISITOS PARA EAD AO VIVO

As aulas serão transmitidas através de uma plataforma digital;

Conexão com a *internet* - banda larga com ou sem fio com pelo menos 1MB de velocidade final, e para acompanhar as aulas ao vivo e para a realização de exercícios;

Computador com configuração mínima: Dual Core 2Ghz ou superior (I3/i5/I7 ou AMD equivalente) com no mínimo 8Gb de RAM;

Os navegadores recomendados são *Internet Explorer* 11+, *Edge* 12+, *Firefox* 27+, *Chrome* 30+ ou *Safari* +7

Permissão de administrador da máquina;

Caso haja determinação legal para aplicação de avaliação presencial, ela será realizada em uma das unidades educacionais da FIA, em São Paulo/SP.

MATRIZ CURRICULAR

O curso será ministrado em dois módulos.

- No primeiro módulo serão apresentadas as tecnologias utilizadas por engenheiros e arquitetos de dados para a resolução de problemas complexos. Neste módulo será criada uma SOLUÇÃO ou PRODUTO baseado em dados que seja resiliente, escalável e confiável utilizando as principais tecnologias de dados destacadas em aula;
- No segundo módulo serão apresentadas aplicações de inteligência artificial utilizadas para a tomada de decisão. Neste módulo serão implementadas as técnicas de *Machine Learning* e *Deep Learning* entregando valor ao projeto proposto;

1º MÓDULO: TECNOLOGIA DE BIG DATA

Este módulo tem como objetivo apresentar, de forma prática, as principais tecnologias utilizadas nas organizações, por engenheiros e arquitetos de dados, para a implementação de projetos de *Big Data* e de Inteligência Artificial.

Introdução

- Conceito de *Big Data*, Inteligência Artificial, *Machine Learning*, Computação Cognitiva e Computação Quântica;
- Panorama para o surgimento do *Big Data*;
- Necessidades e vantagens da utilização das técnicas de *Big Data* e das metodologias de Inteligência Artificial;
- 7 Vs do *Big Data*: Variedade, Veracidade, Valor, Volume, Velocidade, Visualização e Vulnerabilidade;
- O profissional adequado para trabalhar com o *Big Data*: *Data Scientist*, *Data Engineer* e *Data Architect*;
- Cases nacionais e internacionais de *Big Data*.

Cloud Computing

- Fundamentos *Cloud Computing*;
- *Cloud Adoption Framework*;
- *Well architected*;
- Conceitos e casos de uso de aplicações utilizando SaaS (*Software-as-a-Service*), PaaS (*Platform-as-a-Service*) e IaaS (*Infrastructure-as-a-Service*);
- *Finops*;
- Observabilidade.

Programação

- Introdução a linguagem *Python* e suas principais bibliotecas;
- Análise de dados utilizando *Python*;
- Introdução a linguagem SQL;
- Conceitos de DML, DDL e DQL;
- Acesso a dados com linguagem SQL.

Plataformas de Dados

- Criação de plataforma de dados utilizando tecnologias de código aberto;
- *Hadoop*: Arquitetura e conceitos do *framework* pioneiro dos ecossistemas *Big Data*;
- Conceitos e utilização de conjuntos de práticas de automatização de processos DevOps;
- Introdução a *Docker* e *containers*;
- Abordagem de arquitetura em micros serviços independentes;
- EDA - Conceitos da Arquitetura Orientada a Eventos e sua utilização com *Streams* de Eventos;
- MLOPS: Automatização de projetos de *Machine Learning* e suas práticas.

Armazenamento de Dados

- Utilização de dados estruturados, semiestruturados e não estruturados;
- Tipos de arquivos (CSV, JSON, Avro, Parquet, Delta e outros);
- Arquitetura e armazenamento de dados destruído com HDFS;
- Conceito, arquitetura e utilização de padrão S3 em *cloud* e *on premise* com Minio;
- Conceito, arquitetura, modelagem e utilização de banco de dados NoSQL das famílias *Column Store*, *Document* (MongoDB e Elastic), *Key/Value* (Redis);
- Arquitetura de dados relacional, *internal engine* e utilização de bancos de dados relacionais e *Data Warehouse* com MySQL e PostgreSQL.

Ingestão de Dados

- Fontes de dados e coleta de dados;
- *Streaming* de dados com *Kafka Platform*;
- Eventos em banco de dados utilizando CDC (*Change Data Capture*);
- Ingestão em tempo real e fluxo de dados com *Kafka Connect*, *Spark*, *Nifi* e *AirByte*;
- Ingestão *batch*, ETL e ELT utilizando *Spark* e *AirFlow*;
- Orquestração e *deploy* de *pipeline* de ingestão de dados.

Processamento e Análise de Dados

- Conceitos e Arquitetura *Spark*;
- Introdução ao processamento distribuído;
- Análise de dados com *Python* (PySpark) e SQL;
- Análise Exploratória de Dados com *Spark*;
- Processamento de dados em tempo real;
- Análise e cruzamento de dados em modo *batch*;
- Criação de aplicações distribuídas com *Spark*;
- *Spark* na *Cloud* (EMR e *Databricks*);
- Processamento ETL com DBT.

Acesso e Distribuição de Dados

- Conceito de acesso a arquivos com *engine SQL* como *Big Query*, *Athena*, *Synapse*, *Trino*, *Presto* e *Dremio*;
- Arquitetura e acesso a dados com *Hive*;
- Federação de dados e acesso *In Memory* com *Trino/Presto*;
- Entregando dados com API.

Governança

- Catálogo de metadados;
- Qualidade de dados;
- Linhagem;
- Gestão de acesso.

Projeto de Big Data

Será realizado um projeto de *Big Data* com as principais tecnologias utilizadas por engenheiros e arquitetos de dados.

O objetivo deste projeto é aplicar os conhecimentos adquiridos no curso para criar uma solução ou produto baseado em dados que seja resiliente, escalável e confiável utilizando as principais tecnologias de dados destacadas em aula.

Para isso, é proposto desenhar e implementar uma arquitetura de dados com os seguintes pilares:

1. Coleta de Dados:

- Implementar um sistema de ingestão de dados capaz de lidar com diferentes fontes, como bancos de dados, *logs* e *streaming* de dados em tempo real;
- Utilizar ferramentas como Plataforma *Kafka*, *Spark Streaming*, *AirByte*, *Nifi* e *Ariflow* para garantir a escalabilidade e a confiabilidade na ingestão de dados.

2. Armazenamento:

- Projetar e implementar um sistema de armazenamento utilizando tecnologias como HDFS e S3;
- Integrar bases de dados NoSQL, como *Elastic* ou MongoDB, para acesso de dados em baixa latência.

3. Processamento de Dados:

- Aplicar técnicas de processamentos distribuídos utilizando *Apache Spark* e *Python* para realizar análises relevantes ao projeto;
- Implementar *pipelines* de processamento de dados eficientes para transformação e limpeza dos dados.

4. Análise e Visualização:

- Disponibilizar acesso aos dados armazenados e processados com ferramentas de acesso como *Presto*, *Hive*, API, entre outros;
- Desenvolver *dashboards* interativos utilizando *frameworks Metabase* ou *Apache Superset* para fornecer *insights* acionáveis aos usuários finais.

5. Segurança e Gerenciamento de Dados:

- Implementar políticas de segurança para proteger dados sensíveis;
- Utilizar ferramentas de gerenciamento de metadados para rastrear a proveniência dos dados e garantir a conformidade com regulamentações.

6. Entregáveis:

- Documentação técnica detalhada do sistema, abrangendo arquitetura, processos de ETL, segurança e manutenção;
- Código-fonte do sistema, hospedado em um repositório versionado (por exemplo, *GitHub*);
- Apresentação executiva destacando os principais resultados e benefícios obtidos com a implementação da solução.

2º MÓDULO: INTELIGÊNCIA ARTIFICIAL

Este módulo tem como objetivo apresentar os fundamentos e a importância da inteligência artificial para a tomada de decisão.

Inteligência Artificial

- Introdução a inteligência artificial;
- Inteligência Artificial Aplicada aos Negócios;
- Importância da análise de imagem, som e texto.

Análise Exploratória de Dados

- Tipos de variáveis: dados qualitativos e quantitativos;
- Distribuição de frequências;
- Medidas de posição e dispersão;
- Gráficos: *Box Plot* e Histograma;
- Aplicações.

Técnica de Classificação - Regressão Logística

- Importância da regressão logística para a tomada de decisão;
- Tabela de classificação;
- Aplicações.

Técnica de Segmentação - Análise de Agrupamento

- Importância da análise de *cluster* para a tomada de decisão;
- Aplicações.

Regressão Linear Simples e Múltipla

- Importância da regressão linear para a tomada de decisão;
- Aplicações.

Machine Learning

- Aprendizado supervisionado e não supervisionado.
- A importância dos principais algoritmos:
 - *Random Forests* (Florestas Aleatórias);
 - Árvore simples;
 - *CatBoost*;
 - *LGBM (LightGBM)*;
 - *XGBoost (eXtreme Gradient Boosting)*;
 - *SVM (Support Vector Machine)*;
 - *K-means* e outros.

Deep Learning

- Introdução a Redes Neurais;
- Importância da Análise de imagem, som e texto;
- Redes neurais densas;
- Redes recorrentes;
- Redes convolucionais;
- Modelos generativo.

Projeto de Analytics

Neste projeto serão implementadas técnicas de *Machine Learning* e *Deep Learning* entregando valor ao projeto de engenharia de dados proposto no módulo de tecnologias de *big data*.

INFORMAÇÕES

📍 Av. Paulista, 302 – 5º andar,
Bela Vista, São Paulo/SP

✉ labdata@fia.com.br

📞 (11) 98193-2288

🌐 fia.com.br