

Conteúdo	PDFs	Lei Seca	Videoaulas	Questões	Revisão			
					24 h	7 dias	30 dias	90 dias
PROVA DISCURSIVA (MATERIAL - CLIQUE AQUI)								
CONHECIMENTOS BÁSICOS								
LÍNGUA PORTUGUESA								
1. Compreensão de textos.								
2. Ortografia oficial.								
3. Mecanismos de coesão textual.								
4. Significação das palavras.								
5. Emprego de tempos e modos verbais.								
6. Emprego das classes de palavras.								
7. Coordenação e de subordinação.								
8. Emprego dos sinais de pontuação.								
9. Concordância verbal e nominal.								
10. Regência verbal e nominal.								
11. Emprego do sinal indicativo de crase.								
12. Colocação dos pronomes átonos.								
LÍNGUA INGLESA								
1. Compreensão de texto escrito em língua inglesa.								
2. Itens gramaticais relevantes para a compreensão dos conteúdos semânticos.								
CONHECIMENTOS ESPECÍFICOS								
I- MATEMÁTICA								
1. Cálculo Básico: funções; limites; derivadas; derivadas parciais; máximos e mínimos; integrais.								
2. Álgebra Linear: vetores e matrizes; operações com vetores e matrizes; tipos de matrizes; transformações lineares; espaços e subespaços vetoriais de R^n ; sistemas de equações lineares; normas (L1, L2, infinita, p-generalizada, Minkowski e Chebyshev), autovalores e autovetores; decomposição matricial (Cholesky e Singular Value Decomposition (SVD)).								

Conteúdo	PDFs	Lei Seca	Videoaulas	Questões	Revisão			
					24 h	7 dias	30 dias	90 dias
PROVA DISCURSIVA (MATERIAL - CLIQUE AQUI)								
II- PROBABILIDADE E ESTATÍSTICA								
1. Fundamentos de probabilidade: definições básicas de probabilidade; axiomas; probabilidade condicional.								
2. Variáveis aleatórias e distribuições de probabilidades: variáveis aleatórias; funções de probabilidade; principais distribuições discretas e contínuas (Uniforme, Binomial, Normal, Poisson, Bernoulli e Exponencial).								
3. Estatísticas Descritivas: medidas de tendência central (média, mediana e moda); medidas de dispersão (variância, desvio padrão e amplitude); medidas de posição (percentis e quartis).								
4. Teoremas fundamentais da probabilidade: independência de eventos; teorema de Bayes; teorema da probabilidade total; lei dos grandes números; teorema central do limite.								
5. Distribuições amostrais: distribuição amostral da média; distribuição amostral da proporção; distribuição qui-quadrado; distribuição t de Student; distribuição F.								
6. Inferência estatística: estimação pontual e intervalar; intervalos de confiança; testes de hipóteses (formulação, tipos de erros, e poder do teste); testes z e t para médias; testes de proporções; testes qui-quadrado para independência e ajuste de Goodness-of-Fit; teste A/B.								
7. Correlação: correlação e causalidade; correlação de Pearson; correlação de Spearman; correlação parcial.								
III- DADOS E BASES DE DADOS								
1. Conceitos fundamentais de dados, informação e conhecimento; dados qualitativos (nominal, ordinal) e quantitativos (discretos, contínuos); tidy data; formatos de arquivos de dados comuns (txt, csv, xlsx, xml e json).								
2. Introdução a Bases de Dados: o que são bases de dados; metadados.								

Conteúdo	PDFs	Lei Seca	Videoaulas	Questões	Revisão			
					24 h	7 dias	30 dias	90 dias
PROVA DISCURSIVA (MATERIAL - CLIQUE AQUI)								
3. Introdução ao armazenamento de dados: armazenamento de arquivos; principais estruturas de armazenamento de dados analíticos (data warehouse, data mart, data lake, data lakehouse), suas diferenças conceituais e casos de uso; armazenamento na nuvem.								
4. Sistemas Gerenciadores de Base de Dados (SGBDs): definição de SGBD; principais funções; principais tipos de SGBDs (SQL e NoSQL) e suas diferenças; transações e índices; controle de acesso; regras de integridade.								
5. Modelo de dados: modelo de entidade-relacionamento (ER); modelo relacional: tabelas, esquemas, chaves (primária e estrangeira), consultas; dados estruturados, semiestruturados e não estruturados; modelo chave-valor; modelo colunar; modelo orientado a documentos; modelo orientado a grafos.								
6. Big Data: conceito de big data; conceitos gerais sobre técnicas e ferramentas para lidar com grandes volumes de dados (Spark, Hadoop, Parquet, HDFS e MapReduce).								
IV- GESTÃO DE PROJETOS DE CIÊNCIA DE DADOS								
1. Ciclo de vida de projetos de ciência de dados.								
2. Metodologias de gestão de projetos de ciência de dados: CRISP-DM, Microsoft Team Data Science Process (TDSP), OSEMN; princípios de métodos ágeis (Scrum/Kanban).								
V- QUALIDADE E PREPARAÇÃO DE DADOS								
1. Metadados: a sua importância para avaliação da qualidade de dados; linhagem de dados;								
2. Coleta de dados: fontes comuns de dados (internas e externas); técnicas de web scraping.								
3. Problemas comuns de qualidade de dados: valores ausentes; duplicatas; outliers; desbalanceamento; erros de imputação.								
4. Preparação de dados: técnicas de tratamento e limpeza de dados; técnicas detecção de vieses; data profiling.								

Conteúdo	PDFs	Lei Seca	Videoaulas	Questões	Revisão			
					24 h	7 dias	30 dias	90 dias
PROVA DISCURSIVA (MATERIAL - CLIQUE AQUI)								
5. Pré-processamento de dados: técnicas de normalização e padronização; discretização; metodologias de codificação de variáveis categóricas (encoding).								
6. Feature engineering: processos para enriquecimento de dados, com criação e seleção de features relevantes; transformações matemáticas e estatísticas comuns em variáveis.								
7. Divisão de dados: técnicas de amostragem; divisão entre treinamento, validação e teste; abordagens para cross-validation.								
VI- MODELAGEM								
1. Pipeline de treinamento de modelos e suas etapas.								
2. Otimização de hiperparâmetros: grid search; random search; algoritmos de otimização avançados; automl; autotuning; autofeature engineering.								
3. Métricas para avaliação e seleção de modelos: métricas para regressão (MSE; RMSE; MAE; R^2 ; R^2 ajustado); métricas para classificação (accuracy, precision, recall, F1-score e ROC-AUC); análise de matriz de confusão; trade-off entre viés e variância; detecção de overfitting e underfitting.								
4. Técnicas de regularização: lasso; ridge; elastic net; dropout; early stopping; batch normalization.								
5. Dados desbalanceados: técnicas para lidar com dados desbalanceados; oversampling; undersampling; dados sintéticos; ajuste de pesos.								
6. Validação de Modelos: K-fold cross-validation; leave-one-out cross-validation; bootstrap.								
7. Modelagem de IA centrada em dados (data-centric).								
VII- CLASSES DE MODELOS								
1. Redução de dimensionalidade: Principal Component Analysis (PCA); LDA; ICA; T-SNE; uso de autoencoders.								
2. Técnicas de clusterização: K-Means; agrupamento hierárquico; Gaussian Mixture Models; DBSCAN.								

Conteúdo	PDFs	Lei Seca	Videoaulas	Questões	Revisão			
					24 h	7 dias	30 dias	90 dias
PROVA DISCURSIVA (MATERIAL - CLIQUE AQUI)								
3. Técnicas de classificação: Regressão logística; K-Nearest Neighbors (KNN); Suport Vector Machines (SVM); Decision Trees (CART); classificadores Naive-Bayes (Binomial-Beta, Poisson-Gama, Normal-Normal); Florestas Aleatórias (Random Forest).								
4. Introdução à regressão: regressão linear simples e múltipla; hipóteses clássicas, método dos mínimos quadrados, diagnóstico e avaliação de modelos de regressão (F-test, coeficiente de determinação, análise de resíduos e demais), testes de significância, intervalos de confiança, análise ANOVA, modelos não lineares (log-log, lin-log, log-lin e inverso).								
5. Ensembling de modelos: Bagging; boosting (AdaBoost, Gradient Boosting, XGBoost, LightGBM e CatBoost); stacking.								
6. Sistemas de recomendação: Filtragem colaborativa (baseadas em usuários ou itens); filtragem baseada em conteúdo; sistemas híbridos; problemas comuns (cold start, escalabilidade, data sparsity).								
7. Modelos de séries temporais: definição; componentes (tendência, sazonalidade, ciclos e ruído); autocorrelação e autocorrelação parcial; conceito e testes de estacionaridade; cointegração; modelos AR, ARMA e ARIMA; modelos de suavização exponencial; modelos de decomposição; modelos de regressão com variáveis temporais (ARIMAX).								
8. Tópicos em regressão: modelos de dados em painel; GLM; regressão espacial; regressão quantílica; regressão de Poisson; modelos VAR; ECM e GARCH.								
9. Introdução a modelos causais: fundamentos de causalidade estatística, experimentos e quase-experimentos, desenho de descontinuidade de regressão, modelos de variáveis instrumentais, diferenças em diferenças, métodos de pareamento.								
10. Redes neurais: Introdução a Redes Neurais Artificiais (arquitetura, funções de ativação, treinamento, forward pass, backpropagation, loss functions, algoritmos de otimização, épocas, batch size e demais); embeddings; redes profundas (deep learning); Redes Neurais Convolucionais (CNNs) e Recorrentes (RNNs); LSTM; GRU; GAN; modelos multimodais.								

Conteúdo	PDFs	Lei Seca	Videoaulas	Questões	Revisão			
					24 h	7 dias	30 dias	90 dias
PROVA DISCURSIVA (MATERIAL - CLIQUE AQUI)								
VIII- PROCESSAMENTO DE LINGUAGEM NATURAL (NLP)								
1. Técnicas de pré-processamento de texto: limpeza; normalização; remoção de stop words; stemming; lematização e demais.								
2. Representação de texto: N-grams; CBoW; FTD-IDF; word embeddings (Word2Vec, GloVe e demais) e document embeddings (Doc2Vec, BERT, ELMo e demais).								
3. Modelagem de tópicos: latent dirichlet allocation (LDA); non-negative matrix factorization (NMF).								
4. Modelos de linguagem: modelos de linguagem tradicionais; redes neurais recorrentes; redes neurais convolucionais; transformers.								
5. Tarefas básicas em NLP: classificação de texto; análise de sentimento; extração de informação (NER; REL); similaridade textual; sumarização de texto; rotulação de partes do discurso (POS-tagging) e tradução automática.								
IX- PROGRAMAÇÃO E FERRAMENTAS								
1. Linguagem de programação Python: sintaxe básica; operadores; variáveis; estruturas de dados (dataframes, listas, matrizes, dicionários e conjuntos); estruturas de controle de fluxo; funções; escopo; método; paralelização de rotinas; serialização e desserialização.								
2. Bibliotecas Python: Pandas (manipulação; limpeza; transformação e pré-processamento de dado); NumPy (operações de arrays); Matplotlib e Seaborn (visualização de dados); TensorFlow; Keras e PyTorch (redes neurais); Scikit-learn e XGBoost (aprendizado de máquina); NLTK e spaCy (processamento de linguagem natural); huggingface (LLM); PySpark (Big data); Beautiful Soup (web scraping); Streamlit (data apps).								
3. Linguagem SQL (Structured Query Language): conceitos introdutórios; comandos básicos para consultas (inserção, atualização e exclusão de dados) e para análise de dados (como funções de agregação, filtros, joins, subconsultas e demais).								
4. Microsoft Power BI: conexão e importação de dados; modelagem de dados; criação de medidas e colunas calculadas; visualizações e gráficos; interações entre visualizações; criação de relatórios e painéis.								

Conteúdo	PDFs	Lei Seca	Videoaulas	Questões	Revisão			
					24 h	7 dias	30 dias	90 dias
PROVA DISCURSIVA (MATERIAL - CLIQUE AQUI)								
X- VISUALIZAÇÃO; STORYTELLING E COMUNICAÇÃO CORPORATIVA								
1. Principais tipos de visualizações e gráficos: tabela; gráfico de barras; linhas; pizza; dispersão; histograma; área; boxplot; bolhas; radar; mapas cartográficos; mapa de calor.								
2. Visualização de dados: princípios de design de gráficos efetivos; principais conceitos de codificação visual; interatividade; acessibilidade em gráficos.								
3. Dashboards: técnicas para construção de interfaces e layout; abordagens para escolha de designs; organização de elementos visuais e gráficos; seleção de gráficos e visualizações; interatividades e drill-downs; acessibilidade.								
4. Storytelling com dados: construção de narrativas visuais e contextualizações; componentes de um storytelling efetivo.								
XI- GOVERNANÇA E SEGURANÇA DE DADOS								
1. Noções de governança de dados (DMBOK): conceitos e objetivos da governança de dados; principais técnicas de qualidade e integridade de dados; princípios de privacidade e proteção a dados.								
XII- GOVERNANÇA, SEGURANÇA E APLICAÇÃO RESPONSÁVEL DE IA								
1. Noções de governança de IA: conceitos e objetivos da governança de IA; gestão de riscos em IA.								
2. Principais riscos e vulnerabilidades relacionados a IA: viés algorítmico; exposição de dados sensíveis; envenenamento de dados de treinamento; ataques adversariais; ataques de manipulação de modelos; roubo de modelos; ataque de inferência; alucinações.								
3. Aplicação de IA responsável: definição; ética; transparência; justiça e equidade; responsabilização; segurança cibernética; compliance regulatório.								