



# Manual do Projeto Integrador V Projetos em Ciência de Dados I

CST de Ciência de Dados para Negócios

Versão 2.0 – Fevereiro de 2026



## Sumário

Introdução .....	3
Objetivo e Escopo do Projeto Integrador V.....	4
Diretrizes Gerais do Projeto Integrador V .....	6
Extensão Universitária no Projeto Integrador V .....	11
Competências Desenvolvidas e Objetivos de Aprendizagem .....	13
Entregáveis do PI5 .....	16
Etapas Essenciais (Linha do Tempo) .....	19
Sistema de Avaliação .....	25
Ficha de Avaliação — Apresentação Final .....	27
Planejamento das Atividades de Extensão Universitária.....	29
Resumo Executivo — Projeto Integrador V (PI5) • CDN .....	32

# Introdução

O Projeto Integrador constitui o eixo estruturante do Curso Superior de Tecnologia em Ciência de Dados para Negócios da Fatec Votorantim, sendo concebido como um componente curricular articulador entre teoria, prática profissional e atividades de extensão universitária. Ao longo dos seis semestres do curso, os Projetos Integradores promovem a aplicação progressiva dos conhecimentos desenvolvidos nas disciplinas, permitindo que os estudantes enfrentem problemas reais provenientes do setor produtivo e da sociedade.

Nos primeiros semestres, os Projetos Integradores concentram-se na compreensão do contexto organizacional, no entendimento dos processos de negócio e na construção das bases analíticas necessárias à atuação em ciência de dados. À medida que o curso avança, os projetos evoluem em complexidade, passando da análise exploratória e visualização de dados para etapas mais avançadas do pipeline de análise e modelagem.

Nesse contexto, o Projeto Integrador V (PI5) representa um momento de consolidação técnica e profissional da formação do estudante. Após as etapas iniciais desenvolvidas nos Projetos Integradores anteriores — especialmente o diagnóstico do negócio, a análise de dados e a construção de hipóteses analíticas — o PI5 direciona o trabalho para a construção de soluções preditivas completas, incorporando técnicas de Aprendizado de Máquina, Processamento de Linguagem Natural e planejamento de implantação em ambiente real.

O PI5 tem como característica central a integração efetiva entre as disciplinas do quinto semestre e a atuação junto a uma empresa parceira, fortalecendo o caráter extensionista do curso. Por meio dessa interação, os estudantes vivenciam situações reais de desenvolvimento de projetos, incluindo levantamento de requisitos, validação contínua com stakeholders, tomada de decisões técnicas e comunicação de resultados para públicos não especializados.

Diferentemente dos projetos anteriores, o foco do PI5 não está apenas na análise dos dados, mas na proposição de uma solução de ciência de dados tecnicamente estruturada e potencialmente implementável. Assim, os grupos devem desenvolver modelos preditivos, avaliar criticamente seus resultados e elaborar um planejamento detalhado de deploy, preparando o projeto para sua etapa final de implementação, que ocorrerá no Projeto Integrador VI.

Dessa forma, o Projeto Integrador V consolida a transição do estudante de um papel predominantemente analítico para uma atuação próxima à prática profissional em ciência de dados, integrando competências técnicas, visão de negócio, trabalho colaborativo e responsabilidade ética no uso de dados. O resultado esperado é a construção de soluções aplicáveis ao contexto organizacional, capazes de gerar valor real para as empresas parceiras e, simultaneamente, evidenciar a maturidade acadêmica e profissional alcançada pelos estudantes ao longo do curso.

# Objetivo e Escopo do Projeto Integrador V

O **Projeto Integrador V: “Projetos em Ciência de Dados I”** é uma iniciativa interdisciplinar que desafia os estudantes do 5º semestre de Ciência de Dados para Negócios a *propor e desenvolver uma solução completa de ciência de dados para um problema real de negócio*, culminando no planejamento de implantação em ambiente de produção. Esse projeto tem como objetivo aproximar ainda mais os alunos do setor produtivo, consolidando a integração das diversas disciplinas do semestre em um desafio aplicado de alta complexidade.

No PI5, espera-se que cada grupo identifique (ou aprofunde) um **problema de negócio concreto** na empresa parceira e conduza todas as etapas de um pipeline de ciência de dados orientado a soluções: desde o entendimento do problema e preparação dos dados, passando pela *construção de modelos preditivos* com técnicas de Aprendizado de Máquina (e eventualmente NLP, se pertinente), até o desenho de um plano de **deploy** que viabilize a implementação real da solução. O escopo, portanto, abrange tanto a dimensão analítica (modelagem e avaliação técnica) quanto a dimensão estratégica e operacional (como essa solução será utilizada e mantida na empresa).

Em termos práticos, os alunos deverão entregar um produto de ciência de dados funcional e aplicado – por exemplo, um modelo preditivo treinado capaz de gerar insights ou previsões úteis para a empresa – junto com toda a documentação e orientações para sua futura implantação. Caso o grupo esteja dando continuidade ao projeto iniciado no PI3/PI4, o escopo do PI5 será aprofundar a solução previamente delineada, incorporando modelos preditivos e arquitetura de deploy ao que já foi desenvolvido. Se for um projeto novo, o grupo deverá rapidamente alinhar o escopo com a empresa e cobrir etapas iniciais de diagnóstico já nas primeiras semanas, de modo a focar nas etapas de modelagem e deploy pelo restante do semestre. Em ambos os casos, o resultado esperado é *um plano detalhado de solução de ciência de dados pronta para ser implementada* (que efetivamente ocorrerá no PI6).

Esse escopo integrado atende tanto aos objetivos acadêmicos do curso quanto às demandas reais das empresas parceiras, promovendo uma compreensão holística do ciclo de vida de projetos de dados – do problema à solução em produção. Ao envolver análise de dados, modelagem preditiva e considerações de infraestrutura, o PI5 prepara os alunos para os desafios concretos da carreira em Ciência de Dados, onde competências técnicas devem alinhar-se com objetivos de negócio e boas práticas de governança de TI. Em suma, o Projeto Integrador V expande o horizonte dos estudantes, levando-os a arquitetar soluções de dados completas e aplicáveis no mundo real, reforçando a ponte entre conhecimento acadêmico e aplicação prática.

## Disciplinas Âncoras e Integração Curricular

O Projeto Integrador V constitui atividade interdisciplinar estruturante do 5º semestre, sendo desenvolvido em articulação direta com as seguintes disciplinas âncoras:

- Aprendizado de Máquina Aplicado
- Processamento de Linguagem Natural (PLN)
- Análise Preditiva
- Infraestrutura para Ciência de Dados

A nota final do Projeto Integrador V compõe **20% da média final de cada uma das disciplinas do semestre**, conforme diretrizes do Projeto Pedagógico do Curso e deliberações do colegiado.

Dessa forma, o desempenho no PI5 possui caráter transversal e impacta diretamente a avaliação acadêmica global do estudante no semestre.

## Diretrizes Gerais do Projeto Integrador V

Nesta seção, apresentamos as diretrizes oficiais que norteiam a execução do Projeto Integrador V, conforme estabelecido pelo Projeto Pedagógico do Curso de Ciência de Dados para Negócios, incluindo aspectos de parceria com empresa, condução do projeto, ferramentas de suporte, documentação e ética profissional:

- **Formação de grupos e continuidade:** Os grupos de trabalho devem ser definidos **na primeira semana de aula**, sob orientação do(a) professor(a) responsável pelo Projeto Integrador V. Cada grupo pode ter até 6 estudantes, e é altamente recomendável que sejam mantidas as formações do PI anterior (PI4). A continuidade dos mesmos times e do mesmo parceiro empresarial garante um ganho de maturidade e aprofundamento, permitindo ao grupo evoluir a solução sem curva de aprendizado organizacional. Caso algum aluno novo ingresse ou haja necessidade de remanejamento de grupos, isso deve ser alinhado com o professor responsável para preservar o equilíbrio de competências. A sinergia da equipe, construída ao longo dos projetos anteriores, é um ativo importante: facilita a divisão de responsabilidades técnicas (desenvolvimento de modelos, codificação, documentação) e gerenciais (comunicação com a empresa, gestão de cronograma).
- **Empresa parceira e interação contínua:** Cada grupo trabalhará em colaboração com uma empresa parceira real, preferencialmente a mesma definida no PI4, para dar sequência ao projeto. Se uma nova empresa for escolhida, deve-se validar junto ao professor responsável se ela atende aos critérios (segmento adequado, disponibilidade de informações e interesse em colaborar). A interação com a empresa deverá ser *contínua ao longo de todo o semestre*, por meio de visitas técnicas, reuniões (presenciais ou virtuais) e trocas de feedback frequentes. Essa cocriação com o stakeholder garante que a solução proposta permaneça aderente à realidade e às necessidades da organização. É importante agendar, já no início, encontros periódicos (quinzenais, por exemplo) para apresentar progresso, obter validações e redirecionar esforços se necessário. Todos os contatos com representantes da empresa devem ser documentados (atas ou sumários de reunião), servindo tanto como evidência de extensão quanto para registro de requisitos e feedbacks. Os grupos devem estar atentos a possíveis requisitos de *confidencialidade* da empresa: se dados sensíveis forem fornecidos, estabeleça-se um acordo de sigilo e siga-se uma conduta ética exemplar. Lembre-se de que a empresa parceira atua como co-orientadora do projeto, esperando retorno de valor prático – as entregas parciais devem, sempre que possível, ser úteis e aplicáveis para ela.
- **Escopo do projeto e foco do problema de negócio:** O PI5 centra-se na *solução de um problema de negócio utilizando ciência de dados*. É fundamental que o grupo tenha um escopo bem definido e delimitado. Se for continuação do projeto anterior, revise o problema norteador e atualize o escopo conforme os aprendizados do PI4 – por exemplo, incorporando novas perguntas de negócio decorrentes das análises já feitas ou ajustando objetivos conforme feedback da empresa. Se for um projeto novo, dedique as primeiras semanas para compreender o contexto da empresa (mercado, operações, estratégias) e identificar um problema relevante que possa ser endereçado via análise de dados e modelagem preditiva. Em ambos os casos, formule uma pergunta norteadora clara ou um conjunto de objetivos específicos: por exemplo, “*Como prever a demanda de produto X para otimizar o estoque?*” ou “*Qual o risco de churn de clientes (taxa de*

cancelamento) e como reduzi-lo via intervenções personalizadas?”. Um escopo bem definido orientará todas as etapas subsequentes e servirá como referência para evitar “desvios” do projeto. É importante equilibrar ambição e viabilidade: o problema deve ser desafiador o suficiente para exigir técnicas avançadas de ciência de dados, mas ao mesmo tempo factível de ser trabalhado com os dados e recursos disponíveis no intervalo de um semestre. Definir claramente as fronteiras do projeto (o que está incluído e o que está fora do escopo) ajudará na gestão do tempo e na objetividade das análises.

- **Coleta e preparação de dados:** Uma vez definido o problema, o grupo deve planejar e executar a *coleta de dados* necessários. Isso pode envolver: extração de dados internos da empresa (bancos de dados corporativos, planilhas fornecidas, registros de sistemas), coleta de dados públicos externos (bases governamentais, APIs abertas, web scraping de informações relevantes) e até geração de dados via pesquisas ou sensores, conforme o caso. Garantir a *qualidade e suficiência dos dados* é etapa crítica – verifique a representatividade, atualidade e integridade dos dados obtidos. Frequentemente, será necessário *tratar e preparar os dados* antes da análise: limpeza de inconsistências, preenchimento de valores ausentes, transformação de variáveis (normalização, categorização), integração de múltiplas fontes etc. No PI5, espera-se que os alunos lidem com conjuntos de dados potencialmente maiores ou mais complexos do que nos projetos anteriores, possivelmente incluindo dados não estruturados (texto, dados de mídias sociais etc.) em virtude da disciplina de PLN. A infraestrutura de dados também entra em foco: considerem onde os dados serão armazenados e processados (por exemplo, utilizar bancos NoSQL ou serviços em nuvem fornecidos pela empresa/escola, se aplicável). Documentem todo o pipeline de obtenção e preparação dos dados de forma reproduzível – códigos ou notebooks Python para ETL (Extract, Transform, Load) devem ser versionados no GitHub, facilitando revisões e reexecução.
- **Análise exploratória e geração de insights iniciais:** Antes de partir para a modelagem preditiva, o grupo deve realizar uma *Análise Exploratória de Dados (Exploratory Data Analysis – EDA)* aprofundada. Essa etapa, geralmente ocorrendo entre as semanas 5 e 7, envolve examinar distribuições, correlações, tendências temporais, detecção de outliers etc., utilizando técnicas estatísticas e visualizações. O objetivo é compreender os padrões dos dados, validar hipóteses iniciais e até descobrir novos insights que possam refinar o direcionamento do modelo. Por exemplo, identificar quais variáveis parecem mais correlacionadas com o resultado de interesse, se há segmentações naturais nos dados (clusters de clientes, padrões sazonais etc.) ou relações não lineares que requeiram transformação. Todo achado relevante deve ser documentado e, preferencialmente, comunicado à empresa para validação inicial – assim, evita-se prosseguir com suposições equivocadas. Ferramentas de visualização (Matplotlib, Seaborn, Plotly, Power BI etc.) podem ajudar a criar gráficos e dashboards exploratórios. Essa análise também subsidiará escolhas de *features* para os modelos (feature engineering): a partir dos insights, o grupo pode decidir criar variáveis derivadas, combinar atributos, filtrar outliers ou segmentar o problema (por exemplo, modelar certos segmentos de clientes separadamente). A EDA bem feita é base para uma modelagem bem-sucedida, e demonstra domínio de análise de dados, por isso deve constar no relatório com exemplos de gráficos e discussões dos resultados encontrados.
- **Modelagem preditiva (machine learning) e avaliação:** Esta é a etapa central do PI5. Com dados preparados e insights em mãos, os alunos devem selecionar algoritmos de *Machine Learning* adequados ao problema (ex.: regressão linear/múltipla, árvores de decisão, random forest, XGBoost, redes neurais, SVM, modelos de NLP como classificação de texto ou word embeddings etc.) e desenvolver *modelos preditivos*. É recomendada a abordagem iterativa: começar com

modelos mais simples (*baseline*) para ter um ponto de comparação, e gradativamente testar modelos mais complexos ou técnicas avançadas ensinadas nas disciplinas. Por exemplo, para um problema de previsão de séries temporais de demanda, iniciar com uma média móvel ou regressão, e depois avançar para um modelo ARIMA ou Prophet; para um problema de classificação de clientes, talvez começar com regressão logística e evoluir para árvores ou redes neurais conforme necessário. Os modelos devem ser treinados e validados com rigor, utilizando boas práticas como divisão treino/teste (e possivelmente validação cruzada), uso de conjunto de validação, tuning de hiperparâmetros (p.ex., *Grid Search* ou *Random Search*) e verificação de overfitting. *Métricas de avaliação* apropriadas devem ser escolhidas conforme o tipo de problema: MAE, RMSE ou MAPE para regressão; acurácia, precisão, recall, F1, AUC etc. para classificação; métricas específicas se for NLP (por exemplo, BLEU score em tradução etc.). Documente no relatório o desempenho de cada modelo testado, preferencialmente em tabelas ou gráficos comparativos, justificando a escolha do modelo “campeão” que será adotado na solução. Importante: além das métricas quantitativas, considere a *viabilidade prática* do modelo – tempo de treinamento, exigência de recursos computacionais, interpretabilidade e aderência ao que a empresa espera. Por exemplo, um modelo ligeiramente menos preciso porém mais interpretável pode ser preferível em certos cenários corporativos. Sempre que possível, envolva a empresa nessas discussões de trade-off, explicando em termos não técnicos os resultados para obter feedback (isso também demonstra habilidade de tradução de conhecimento técnico). Vale ressaltar a necessidade de aplicar conceitos sobre *Ética e IA*: avaliar se o modelo possui vieses, se há riscos de uso indevido das previsões e pensar em mitigação (ex.: remover atributos sensíveis que possam causar discriminação inadvertida). Ao final da etapa de modelagem, espera-se que o grupo tenha um modelo preditivo treinado e validado, pronto para ser utilizado na resolução do problema proposto.

- **Integração de NLP (se aplicável):** Dada a disciplina de *Processamento de Linguagem Natural (PLN)* integrada neste semestre, é incentivado que projetos cujo problema envolva dados textuais aproveitem técnicas de NLP. Por exemplo, se a empresa deseja analisar sentimento de clientes em redes sociais ou classificar documentos (e-mails, avaliações) automaticamente, os alunos podem aplicar modelos de classificação de texto, análise de sentimento, extração de palavras-chave etc. Isso implicaria em etapas específicas no pipeline: coleta de textos, limpeza (remoção de stopwords, tokenização), eventualmente vetorização (TF-IDF, *word embeddings* com Word2Vec/GloVe ou transformers), e então modelagem (p. ex., Naive Bayes, modelos baseados em redes neurais como BERT etc.). Integrar NLP ao projeto não é obrigatório para todos os grupos – somente se o escopo definido tiver componente textual significativo – mas agrega valor multidisciplinar ao PI5 e permite demonstrar competência numa área em alta no mercado. Caso utilizem, devem documentar os procedimentos de PLN no relatório (pré-processamento de linguagem, escolha de modelos, exemplos de resultados qualitativos). A decisão de incluir ou não técnicas de NLP deve ser guiada pelo fit ao problema de negócio: use-as se efetivamente ajudar a resolver a questão proposta ou gerar insights adicionais para a empresa.
- **Planejamento de deploy e infraestrutura de dados:** Um diferencial do PI5 é ir além da modelagem e delinear *como a solução de ciência de dados será implantada na prática* na empresa. Os grupos devem elaborar um plano de deploy detalhado, que aborde: arquitetura de sistema para colocar o modelo em produção, tecnologias envolvidas, cronograma estimado de implementação e considerações de manutenção. Por exemplo, definir se o modelo será exposto via uma API REST em um serviço web (Flask/FastAPI ou similar), se será integrado a um sistema já existente da

empresa (CRM, ERP), ou implantado na nuvem (AWS, Azure, GCP) usando contêineres Docker etc. Considere também *pipeline de atualização*: como novos dados serão incorporados e o modelo será re-treinado periodicamente (MLOps básico), e *monitoramento* do desempenho do modelo em produção (acompanhamento de métricas de acurácia ao longo do tempo, detecção de *drift*). Como parte do plano, recomenda-se incluir um diagrama de arquitetura ilustrando componentes e fluxo de dados (por ex.: fonte de dados → etapa de ETL → modelo em API → front-end ou sistema consumidor). Detalhe quais ferramentas de Big Data ou infraestrutura serão utilizadas conforme aprendido na disciplina específica – por exemplo, o uso de um banco de dados NoSQL para armazenar os resultados, ou de um serviço de fila para processar previsões em lote etc., caso relevantes. Importante: o *deploy efetivo não será realizado no PI5* (pois isso acontecerá no PI6), mas o plano deve ser realista e alinhado com as capacidades da empresa parceira (não adianta propor uso de tecnologia que a empresa não tem como suportar). Esse planejamento demonstra visão prática e prepara terreno para o próximo semestre. Deve constar no relatório um capítulo ou seção específica sobre “Planejamento de Implantação”, descrevendo tudo isso de forma clara, e o guia técnico entregue também deve refletir essas instruções de utilização da solução em ambiente real.

- **Gestão ágil e entregas parciais:** A condução do PI5 deve seguir princípios de *metodologias ágeis de projeto*. Assim como no PI4, o uso de um quadro kanban (Trello ou similar) é obrigatório, refletindo o backlog de tarefas, itens em andamento e concluídos. Recomenda-se que o trabalho seja dividido em *sprints curtos de 2 a 3 semanas*, cada qual resultando em *entregas técnicas parciais* que possam ser validadas pela empresa. Por exemplo, ao final de um sprint, o grupo pode entregar um *dataset* tratado e aprovado, em outro sprint uma versão inicial do modelo ou um dashboard de resultados preliminares, e assim por diante. Essas microentregas devem ser combinadas com a empresa (de forma que alguém do lado deles dê um *feedback* ou aceite formalmente, servindo como evidência de validação contínua). A utilização de sprints e entregas frequentes ajuda a manter o projeto no rumo certo, evitando surpresas no final, e permite incorporar feedback de forma incremental. Os alunos devem aproveitar as ferramentas ágeis (cartões, etiquetas de prioridade, definição de *sprints* no Trello ou similar) para organizar essas iterações. Lembre-se de atualizar o professor(a)-orientador(a) sobre o andamento regularmente; muitas vezes serão agendadas *sessões de acompanhamento quinzenais* na disciplina de PI para discussão de obstáculos e próximos passos. No repositório GitHub, é interessante usar a seção de *Issues* ou Wiki para registrar decisões tomadas em cada iteração e manter um log de progresso.
- **Documentação centralizada e colaborativa:** Uma diretriz importante do PI5 é que *toda a documentação do projeto seja mantida de forma centralizada*, preferencialmente no GitHub do grupo. Cada grupo deve criar (se ainda não o fez) ou continuar a usar um repositório próprio para o PI, que servirá como portfólio vivo do projeto. Nele devem constar: o código-fonte completo (scripts Python, notebooks Jupyter, arquivos de configuração), o relatório técnico em desenvolvimento (por exemplo, à medida que escrevem partes do relatório, podem manter versões em formato texto/Markdown ou Word no repositório), dados ou amostras de dados tratados que possam ser disponibilizados (respeitando confidencialidade), o guia técnico, e quaisquer outros artefatos (slides, atas de reunião etc.). O uso do GitHub traz benefícios pedagógicos e práticos: permite controle de versão (é possível recuperar versões anteriores do trabalho), facilita a colaboração simultânea (vários membros editando seções diferentes sem conflito), e deixa transparente a contribuição de cada membro (histórico de *commits*). O professor responsável pelo PI 5 pode acompanhar periodicamente o repositório para dar orientações (assim como os demais

professores das disciplinas integradas ao PI deste quinto semestre). Dicas: mantenha a estrutura do repositório organizada (pastas para *data*, *src*, *docs* etc.), escreva mensagens de *commit* claras e significativas, e não deixe para subir um grande volume de alterações apenas no final – faça *commits* incrementais ao longo do semestre. Além do GitHub, incentive-se o uso de recursos como GitHub Projects (para gestão de tarefas integrado, em alternativa ao Trello) e GitHub Releases para marcar versões importantes (por exemplo, “Versão Preliminar do Relatório – v0.9”). A documentação de código deve seguir boas práticas: comentários explicando partes complexas, *readme* no repositório explicando como reproduzir o projeto etc. Ao final, espera-se que o repositório GitHub funcione como um repositório acadêmico e profissional do projeto, que poderá inclusive ser compartilhado (respeitando NDA, se houver) como portfólio em entrevistas de estágio/emprego.

- **Normas de formatação e escrita acadêmica:** O relatório técnico final do PI5 será um documento acadêmico extenso, por isso deve obedecer aos padrões formais de *formatação e metodologia científica*. Desde o início, os alunos devem se preocupar em seguir as normas ABNT quanto à estrutura (capa, folha de rosto, sumário, introdução, desenvolvimento, conclusão, referências, apêndices/anexos), formatação de texto (fonte Times 12, espaçamento 1,5, margens etc., conforme guia institucional) e estilo de escrita. Recomenda-se consultar o **Manual de TCC/relatórios da Fatec Votorantim** para orientações padronizadas. Toda figura, tabela ou gráfico inserido deve ter legenda e fonte. As referências bibliográficas devem ser listadas ao final em ordem alfabética, no formato autor-data conforme ABNT (sobrenome em maiúsculas, título em negrito, local, editora, ano etc.). Durante a redação, usar citações diretas ou indiretas com o devido crédito (autor, ano) sempre que for utilizada uma informação de fonte externa – evitar terminantemente o plágio. A escrita deve ter tom impessoal e objetivo, privilegiando voz ativa e evitando a primeira ou terceira pessoa (“nós fizemos” pode ser substituído por “foi feito”). É importante também coesão e coerência: o texto do relatório deve “contar a história do projeto” de forma lógica, desde a apresentação do problema de negócio, passando pela metodologia de ciência de dados adotada, até os resultados obtidos e proposta de deploy. Recomenda-se finalizar o relatório com uma conclusão que retome os objetivos e discuta brevemente limitações do trabalho e sugestões de trabalhos futuros (por exemplo, pontos que poderão ser implementados no PI6 ou estudos adicionais que ficaram fora do escopo). Por fim, anexos podem ser adicionados para conteúdos suplementares, como *registros de reuniões* com a empresa, documentos de validação assinados, trechos extensos de código ou tabelas de dados brutas. Esses anexos servem para evidenciar o processo e não contam no limite de páginas caso haja. Lembre-se: seguir as normas e produzir um texto bem escrito não é apenas uma formalidade acadêmica, mas parte da avaliação e um reflexo de profissionalismo.

Em resumo, essas diretrizes gerais estabelecem como o projeto deve ser conduzido. Manter um escopo claro e relevante, planejar bem as etapas, utilizar as ferramentas e metodologias ensinadas, respeitar padrões éticos e acadêmicos, e engajar a empresa parceira durante todo o processo são fatores críticos para o sucesso do PI5. Seguir essas orientações contribuirá para um projeto robusto, bem organizado e com alto valor agregado, tanto do ponto de vista educacional quanto para a empresa envolvida.

## Extensão Universitária no Projeto Integrador V

O Projeto Integrador V incorpora, de forma estruturada, atividades de extensão universitária, conforme as diretrizes nacionais para a curricularização da extensão no ensino superior e as orientações institucionais do Centro Paula Souza. Nesse contexto, a extensão é compreendida como um processo educativo, interdisciplinar e transformador, que promove a interação direta entre a instituição de ensino, o setor produtivo e a sociedade, possibilitando a aplicação do conhecimento acadêmico na resolução de problemas reais.

No âmbito do PI5, a extensão ocorre por meio da colaboração contínua entre os estudantes e uma empresa parceira externa, que atua como ambiente real de aplicação das soluções de ciência de dados desenvolvidas ao longo do semestre. Essa interação permite que os alunos compreendam demandas concretas do mercado, vivenciem desafios profissionais autênticos e proponham soluções baseadas em dados que possam gerar impacto organizacional e social.

Diferentemente de atividades exclusivamente acadêmicas, a extensão pressupõe troca de conhecimento em via dupla. Assim, ao mesmo tempo em que os estudantes aplicam técnicas de análise e modelagem de dados aprendidas no curso, também assimilam conhecimentos práticos sobre processos organizacionais, cultura empresarial, limitações operacionais e necessidades estratégicas das organizações parceiras. Esse processo contribui para a formação de profissionais mais conscientes, críticos e preparados para atuar em contextos reais.

No Projeto Integrador V, as atividades extensionistas estão diretamente relacionadas às etapas avançadas do pipeline de ciência de dados, incluindo a validação contínua das análises junto à empresa, a discussão dos resultados obtidos, a adaptação das soluções às necessidades do parceiro e a elaboração de um plano de implantação aplicável ao ambiente organizacional. Dessa forma, o projeto ultrapassa o caráter simulativo e aproxima-se de uma experiência profissional supervisionada.

Para caracterização das atividades de extensão, é obrigatória a comprovação da interação efetiva com a empresa parceira ao longo do semestre. Essa comprovação deverá ocorrer por meio de evidências documentais, tais como:

- registros de reuniões técnicas (atas ou relatórios sintéticos);
- validações formais das entregas parciais;
- comunicações institucionais com representantes da empresa;
- devolutivas técnicas realizadas pelos estudantes;
- declarações ou manifestações formais do parceiro externo.

Essas evidências devem demonstrar a participação ativa da organização no desenvolvimento do projeto e a aplicação prática dos resultados produzidos.

As atividades extensionistas desenvolvidas no PI5 contribuem diretamente para o cumprimento da carga horária de extensão prevista no curso, reforçando o compromisso institucional com a formação cidadã, a inovação aplicada e o desenvolvimento regional. Ao atuar em problemas reais, os estudantes ampliam sua compreensão sobre o papel social da ciência de dados, reconhecendo a importância do uso responsável das informações e da geração de valor por meio da tecnologia.

Assim, a extensão universitária no Projeto Integrador V consolida-se como elemento essencial do processo formativo, integrando ensino, prática profissional e impacto social, ao mesmo tempo em que prepara os estudantes para a etapa final do ciclo de projetos integradores, representada pelo Projeto Integrador VI.

# Competências Desenvolvidas e Objetivos de Aprendizagem

O Projeto Integrador V foi concebido para desenvolver nos estudantes um conjunto de *competências profissionais e socioemocionais* avançadas, bem como alcançar determinados *objetivos de aprendizagem* específicos no contexto de projetos de ciência de dados aplicada. A seguir, listamos essas competências e objetivos, com uma breve explicação de cada um:

## Competências Desenvolvidas

- **Formulação de soluções completas em ciência de dados:** Capacidade de *entender um problema de negócio e traduzi-lo em uma solução de dados end-to-end*, abrangendo desde a análise exploratória inicial até a concepção de um modelo preditivo e seu plano de implementação. Envolve pensamento sistêmico para conectar objetivos estratégicos da empresa com técnicas de ciência de dados e propor soluções viáveis e impactantes.
- **Desenvolvimento de modelos preditivos (Machine Learning):** Habilidade de aplicar algoritmos de aprendizado de máquina para *construir, ajustar e validar modelos preditivos* eficazes. Isso inclui desde modelos estatísticos básicos até técnicas avançadas aprendidas na disciplina de Aprendizado de Máquina Aplicado, contemplando tratamento de dados, seleção de variáveis, tuning de hiperparâmetros e avaliação com métricas adequadas. A competência envolve também senso crítico para comparar modelos distintos e justificar a escolha do mais apropriado ao caso.
- **Aplicação de Processamento de Linguagem Natural (PLN):** Competência de manipular e extrair informação de *dados não-estruturados (textos)*, desenvolvendo soluções baseadas em linguagem natural quando pertinente. Engloba conhecimentos de pré-processamento textual, técnicas de vetorização (bag-of-words, TF-IDF, embeddings) e uso de modelos de NLP (como classificadores de texto, análise de sentimento etc.) para agregar valor ao projeto. Mesmo que nem todo projeto use NLP, espera-se que o aluno tenha desenvolvido noções de quando e como aplicar essas técnicas no contexto de negócios.
- **Infraestrutura de dados e planejamento de deploy:** Capacidade de compreender os requisitos de infraestrutura necessários para colocar em produção uma solução de ciência de dados. Isso inclui conhecimentos de *arquitetura de sistemas*, bancos de dados, computação em nuvem e conceitos de Big Data adquiridos na disciplina de Infraestrutura para CDN. Os alunos desenvolvem a habilidade de *planejar um deploy* – por exemplo, decidir entre deploy local vs. em nuvem, uso de contêineres, orquestração, pipelines de CI/CD – e antecipar desafios de manutenção, escalabilidade e segurança da solução proposta.
- **Gerenciamento de projetos e metodologias ágeis:** Habilidade de *gerir um projeto de ciência de dados complexo* ao longo de um semestre, definindo metas, prazos e responsabilidades. No PI5, isso se traduz em organizar o trabalho em sprints, acompanhar tarefas via Trello (ou similar), conduzir reuniões produtivas com stakeholders e adaptar o plano conforme necessidades (agilidade). Também implica em desenvolver resiliência e resolução de problemas diante de obstáculos (por exemplo, dados indisponíveis ou resultados iniciais insatisfatórios), mantendo a equipe focada no objetivo final.
- **Colaboração e comunicação efetiva:** Competência de trabalhar de forma colaborativa em equipe multidisciplinar, comunicando ideias e resultados de maneira *clara e eficaz* tanto para

públicos técnicos quanto não técnicos. Os alunos aperfeiçoam a divisão de tarefas conforme especialidades, aprendem a integrar contribuições individuais em um todo coerente e praticam a comunicação frequente com a empresa parceira (incluindo apresentações parciais, redação de e-mails de atualização, etc.). Além disso, desenvolvem a capacidade de *comunicar resultados de ciência de dados* – como explicar os insights de um modelo ou justificar decisões técnicas – de forma compreensível para gestores ou clientes, refinando habilidades de visualização de dados e storytelling.

- **Ética, responsabilidade e profissionalismo em dados:** Desenvolvimento de uma postura ética sólida ao conduzir análises e construir modelos. Isso inclui respeito à privacidade e confidencialidade dos dados, consciência sobre o impacto das soluções de IA na sociedade (evitando vieses e usos indevidos), e compromisso com integridade acadêmica (sem plágios, com referências adequadas). Também abrange profissionalismo no relacionamento com a empresa (cumprir horários, formalidade na comunicação, confiabilidade nas entregas). Em síntese, o aluno torna-se apto a atuar de forma ética e responsável em projetos de dados no mundo real, ciente de seu papel como futuro profissional.
- **Documentação técnica e relato científico:** Competência de produzir *documentação de qualidade* tanto para usuários/técnicos (guia de uso, código bem comentado) quanto textos acadêmicos (relatório ABNT). Os estudantes aprimoram a habilidade de escrever relatórios estruturados, descrevendo metodologias e resultados com clareza e sustentação teórica, bem como de elaborar manuais práticos que orientem a implementação/uso da solução. Essa competência garante que o conhecimento gerado no projeto seja registrado e transmitido de forma eficiente, atendendo aos padrões exigidos pela comunidade acadêmica e pelas empresas.

## Objetivos de Aprendizagem

Ao final do Projeto Integrador V, os alunos deverão ser capazes de atingir os seguintes *objetivos de aprendizagem*, demonstrando na prática o que foi assimilado durante a execução do projeto:

1. **Identificar e delimitar problemas de negócio complexos para solução com ciência de dados:** Ser capaz de compreender demandas reais da empresa, levantando requisitos e traduzindo-os em perguntas concretas que podem ser respondidas por análises de dados ou modelos preditivos. Esse objetivo envolve realizar reuniões de levantamento com stakeholders e produzir um *escopo bem definido* do projeto, alinhado com as estratégias da organização.
2. **Integrar e aplicar conhecimentos multidisciplinares na elaboração de uma solução de dados:** Utilizar de forma sinérgica os conceitos aprendidos nas disciplinas do semestre – algoritmos de machine learning, técnicas de NLP, métodos de análise preditiva e fundamentos de infraestrutura – para *desenhar um pipeline de ciência de dados completo*. Na prática, o aluno deve demonstrar que sabe combinar técnicas (por exemplo, usar modelagem preditiva em conjunto com análise exploratória e conhecimento de domínio) para resolver o problema proposto de ponta a ponta.
3. **Coletar, limpar e analisar dados de diversas fontes de forma autônoma:** Demonstrar proficiência em *obtenção e preparação de dados*, identificando fontes adequadas (internas ou externas), realizando ETL e garantindo a qualidade dos dados para a modelagem. Espera-se que o aluno consiga lidar com desafios como dados faltantes, ruídos, diferentes formatos e volume considerável de informações, documentando o processo e justificando as decisões de preparação de dados tomadas.

4. **Desenvolver modelos preditivos e avaliar seu desempenho criticamente:** Construir modelos de machine learning adequados ao problema, configurar experimentos de treinamento/validação e *interpretar os resultados obtidos*. O aluno deve ser capaz de calcular e explicar métricas de desempenho, entender as limitações do modelo e iterar em melhorias. Além disso, deve saber comunicar as conclusões desses modelos – por exemplo, quais fatores foram mais importantes na predição – e propor ações ou recomendações baseadas nas predições (conectando o resultado técnico ao contexto de negócio).
5. **Utilizar ferramentas de NLP para extrair insights de dados textuais (quando aplicável):** Se o projeto envolver dados textuais, aplicar corretamente técnicas de PLN para processá-los e analisá-los. O objetivo é que o estudante consiga, por exemplo, extrair sentimentos de avaliações de clientes ou classificar documentos, integrando esses insights textuais ao restante da análise de dados do projeto. Atingir esse objetivo demonstra familiaridade com um tipo de dado frequentemente encontrado em ambientes de big data e complementa a formação em ciência de dados.
6. **Elaborar um plano de deploy para implementação da solução em ambiente de produção:** Apresentar um planejamento estruturado de como o modelo/solução será implementado na prática. O aluno deve ser capaz de *descrever uma arquitetura de implementação*, identificando componentes (servidores, bancos, aplicações) e tecnologias que seriam usadas, e planejar procedimentos de implantação, testes, monitoramento e atualização do modelo. Mesmo sem realizar o deploy no PI5, o estudante deve mostrar que sabe planejar essa fase e está preparado para executá-la (no PI6), considerando aspectos de viabilidade técnica e operacional na empresa.
7. **Documentar e comunicar efetivamente todo o projeto:** Ser capaz de produzir os entregáveis documentais exigidos (relatório técnico e manual de uso) com qualidade, e de apresentar oralmente o projeto de forma clara, objetiva e dentro do tempo. Ao atingir este objetivo, o aluno demonstra aptidão em escrita técnica seguindo normas (ABNT) e em design de apresentações profissionais, além de oratória e domínio do conteúdo para responder questões da banca. Também envolve a habilidade de sintetizar pontos-chave do projeto para diferentes audiências (por exemplo, enfatizando valor de negócio para gestores na empresa e metodologia para avaliadores acadêmicos).
8. **Trabalhar em equipe utilizando metodologias e ferramentas profissionais:** Colaborar eficientemente em grupo, fazendo uso de ferramentas modernas como GitHub (para versionamento/código) e Trello (para gestão ágil) ao longo do projeto. Esse objetivo de aprendizagem se reflete na capacidade do aluno em *atuar profissionalmente* em um ambiente de projeto real: dividindo tarefas conforme as habilidades, integrando as partes entregues pelos colegas, resolvendo conflitos de agenda ou opinião de forma madura, e cumprindo prazos e compromissos. Ao final, espera-se que o grupo funcione como uma pequena equipe de dados dentro da empresa, evidenciando prontidão para o mercado de trabalho.

Essas competências e objetivos balizam todo o desenvolvimento do PI5. Em suma, espera-se que ao término do projeto o aluno tenha consolidado sua capacidade de conduzir um projeto completo de ciência de dados em ambiente corporativo, desde a identificação do problema até a proposta de implantação de um modelo preditivo, integrando conhecimentos técnicos e de negócio. Além disso, que tenha aprimorado habilidades socioemocionais de trabalho em equipe, comunicação e pensamento crítico. O alcance desses objetivos preparará os formandos para desafios reais na carreira, unindo teoria e prática em prol de resultados concretos.

# Entregáveis do PI5

O Projeto Integrador V prevê um conjunto de entregas formais que refletem tanto a dimensão acadêmica (relatórios, análises, fundamentação teórica) quanto a dimensão técnica e extensionista (produto funcional de ciência de dados e devolutivas à empresa). Para garantir clareza, todos os entregáveis esperados estão explicitados a seguir:

- 1. Relatório Técnico Final (formato ABNT):** Documento acadêmico que descreve todo o projeto de forma estruturada e detalhada. Deve conter seções como: **Introdução** (apresentando o contexto da empresa, o problema de negócio e objetivos do projeto), **Referencial Teórico/Fundamentação** (conceitos de ciência de dados, métodos de ML/NLP e ferramentas empregadas, com as devidas citações), **Metodologia** (descrição do pipeline adotado, fontes de dados, critérios de avaliação, etc.), **Análises e Resultados** (achados da análise exploratória, desempenho dos modelos preditivos com tabelas/gráficos, discussões sobre os resultados), **Planejamento de Deploy** (detalhando a arquitetura e estratégias de implementação futura), **Conclusão** (retomando os objetivos e trazendo considerações finais sobre o projeto, incluindo benefícios esperados e possíveis próximos passos no PI6), **Referências** bibliográficas e **Anexos**. O relatório deve seguir as normas ABNT de formatação e citação, conforme orientado anteriormente, e ser escrito de forma coesa e gramaticalmente correta. A expectativa é de um documento robusto, que demonstre rigor acadêmico e aprofunde a solução desenvolvida.
- 2. Apresentação oral (20–30 minutos):** Exposição presencial dos resultados do projeto perante uma banca avaliadora (composta por professores do curso e possivelmente representantes da empresa parceira). A apresentação deve utilizar slides no *template institucional* da Fatec e ter duração aproximada de 20 a 30 minutos por grupo, seguida de sessão de perguntas e respostas. O conteúdo da apresentação deve resumir os principais pontos do relatório: contextualização do problema, abordagem de ciência de dados adotada, principais resultados (com visualizações ilustrativas) e a demonstração do produto funcional (por exemplo, mostrar o modelo fazendo uma previsão ao vivo ou exibindo o dashboard de insights gerado). É importante também abordar, de forma sintética, o plano de deploy – para evidenciar a aplicabilidade futura. Os alunos serão avaliados pela clareza e organização da apresentação, domínio do conteúdo, qualidade visual dos slides e postura profissional durante a exposição. Ressalta-se que a apresentação final é a “vitrine” do trabalho de um semestre, portanto o grupo deve ensaiar previamente e buscar tornar a narrativa envolvente e alinhada ao público (acadêmicos e gestores da empresa).
- 3. Produto Técnico Funcional:** Consiste na *solução de ciência de dados em si*, ou seja, os artefatos computacionais desenvolvidos que materializam o projeto. Inclui tipicamente: o **código-fonte completo** do pipeline de análise/modelagem (scripts Python, notebooks Jupyter, etc.), os **modelos preditivos treinados** (arquivos de modelo ou serializações, como arquivos .pkl ou equivalentes), eventuais **dashboards ou visualizações interativas** produzidos (por exemplo, dashboard em Power BI ou aplicações web simples demonstrativas) e configurações usadas. No PI5, espera-se que o produto técnico inclua ao menos um *protótipo funcional* que possa ser entregue para a empresa parceira testar/utilizar. Por exemplo: um modelo de previsão que rode em um notebook permitindo gerar previsões com novos dados de entrada, ou uma pequena aplicação demonstrativa (não necessariamente com interface complexa, mas

talvez um script de linha de comando ou API local) ilustrando o funcionamento do modelo. Caso o projeto envolva um dashboard interativo, esse dashboard deve estar navegável (mesmo que offline) para a banca/empresa inspecionar. Todos os componentes do produto devem ser entregues de forma reprodutível, preferencialmente via GitHub (ou mídias digitais se aplicável), acompanhados de instruções claras para execução no *guia de uso*. O produto funcional será avaliado quanto à completude (se implementa o que foi proposto), qualidade técnica (boas práticas de código, organização, documentação interna) e utilidade para a empresa. No contexto do PI5, “funcional” significa que, ainda que não esteja implantado em produção, o produto *funciona em ambiente de teste/laboratório*, demonstrando concretamente a solução.

4. **Guia de Uso (Manual Técnico da Solução):** Documento complementar de caráter técnico-operacional, destinado a permitir que a empresa (ou qualquer interessado) consiga instalar, executar e manter a solução desenvolvida de forma autônoma. Deve conter um passo a passo detalhado, incluindo requisitos de ambiente (versão do Python, bibliotecas necessárias – idealmente fornecendo um requirements.txt ou ambiente Conda), instruções de instalação de softwares ou configurações (por exemplo, “*como configurar as credenciais da API X*”, se pertinente), como executar o código ou carregar o modelo para realizar previsões, e exemplos de uso. Caso um deploy em nuvem tenha sido planejado, o guia pode conter orientações iniciais de como reproduzir o ambiente (p.ex., scripts de infraestrutura como Terraform ou Dockerfiles, se os alunos desenvolveram). O guia de uso também deve abranger como interpretar os resultados gerados pela ferramenta – por exemplo, explicando as colunas de saída de uma previsão, ou como navegar no dashboard produzido. Pense nesse documento como aquele que você entregaria a um usuário técnico da empresa (ex.: um analista de TI ou de negócio) para que ele consiga dar continuidade no uso da solução após o término do projeto acadêmico. Diferentemente do relatório (que tem um tom mais teórico e de avaliação acadêmica), o guia de uso é prático e didático, podendo incluir capturas de tela, exemplos de comando etc. Sempre que possível, valide esse documento com alguém da empresa para garantir que está claro e útil.
5. **Evidências de processo:** Coletânea de materiais que comprovem o engajamento do grupo e o processo de desenvolvimento ao longo do semestre. Inclui, entre outros: as **atas ou resumos de reuniões** realizadas com a empresa (contendo data, participantes e principais pontos discutidos/decisões tomadas, preferencialmente assinadas ou validadas pelo representante da empresa), o histórico de **commits no repositório GitHub** (que será analisado pelos avaliadores para verificar contribuição de cada membro e evolução do trabalho), e capturas ou relatórios do **quadro de tarefas (Trello)** mostrando o andamento e fechamento de atividades. Adicionalmente, podem ser anexados comprovantes de entregas parciais feitas à empresa (por exemplo, um e-mail de um gestor confirmando recebimento de uma análise preliminar, ou um formulário de feedback preenchido pela empresa em meio de projeto). Essas evidências de processo podem constar em anexo no relatório ou ser entregues separadamente na plataforma da disciplina, conforme orientação dos professores. O importante é demonstrar transparência e diligência: grupos que mantiveram documentação consistente do processo, com interação real com a empresa, atendem ao caráter extensionista do PI e tendem a ter melhor avaliação.
6. **Validação contínua pela empresa:** Embora parte das evidências de processo, destaca-se como entregável a obtenção de *validações formais da empresa parceira* ao longo do projeto. Isso pode ser materializado em formulários ou documentos assinados pela empresa atestando

cada entrega parcial importante. Por exemplo: assinatura de um termo confirmando que “o conjunto de dados tratado entregue em tal data atende aos requisitos”, ou “o dashboard preliminar apresentado em tal reunião foi recebido e considerado relevante pelo setor X”. No mínimo, espera-se uma **carta ou declaração da empresa ao final do projeto**, confirmando a participação do grupo no projeto e pontuando brevemente a utilidade do trabalho para a organização. Essa validação contínua evidencia que houve alinhamento e feedback real do parceiro externo em todas as etapas críticas. Para o grupo, também serve como métrica de sucesso: se a empresa endossa o que foi feito, é sinal de que o projeto está no caminho certo. Os documentos de validação deverão ser entregues anexos ao relatório final.

Todos esses entregáveis devem ser produzidos com qualidade e entregues nos prazos estipulados no cronograma. Notamos que, no PI5, há um aumento na quantidade e complexidade das entregas em relação aos PIs iniciais – isso reflete o caráter mais avançado e aplicado do projeto. A disciplina de PI5 funcionará, portanto, como um **gerenciador dessas entregas**, orientando os grupos e garantindo que nenhuma parte essencial fique negligenciada. A seguir, detalhamos as **etapas sugeridas e o cronograma** para realização de todas as atividades necessárias à conclusão bem-sucedida do projeto.

Recomenda-se que as evidências sejam organizadas cronologicamente e nomeadas conforme padrão:

**AAAAMMDD\_tipo\_descricao.extensão**

# Etapas Essenciais (Linha do Tempo)

Para orientar o desenvolvimento do projeto ao longo do semestre, apresenta-se uma divisão em etapas essenciais, alinhadas às semanas letivas (considerando ~15 semanas efetivas de aula). Cada etapa corresponde a um conjunto de atividades e *milestones* que conduzem progressivamente das fases iniciais de planejamento até a entrega final e defesa do PI5. A tabela abaixo resume as etapas e um *cronograma sugerido* de 15 semanas, indicando as principais atividades e entregas em cada uma:

- Semana 1 – Introdução ao PI5 e Alinhamento Inicial:** Apresentação da disciplina de PI5 pelos professores, recapitulando o encadeamento de projetos integradores e os objetivos específicos do PI5. Integração com as disciplinas do semestre (breve exposição de como Aprendizado de Máquina, PLN, Análise Preditiva e Infraestrutura contribuirão para o projeto). Formação ou confirmação dos grupos de trabalho. Para grupos já existentes do PI4, confirmar a continuidade e lembrar o problema de negócio em foco; para novos grupos/projetos, esta semana serve para discutir ideias de possíveis empresas e problemas. **Entrega:** nome dos integrantes de cada grupo e, se possível, indicação preliminar se continuarão com a mesma empresa ou se há intenção de nova empresa (a confirmação oficial da empresa ocorre na semana 2). Apresentação do cronograma e critérios de avaliação do PI5 aos alunos, esclarecendo expectativas. (*Dica: já criar o repositório GitHub do grupo e um quadro no Trello nesta semana, incluindo um README inicial com resumo do projeto e primeiros cartões de tarefas.*)
- Semana 2 – Definição da Empresa e Escopo do Projeto:** Cada grupo define e informa oficialmente qual será a empresa parceira e qual o problema de negócio a ser tratado. Para continuidade, confirma-se a aprovação da empresa em prosseguir com o projeto e discute-se se haverá algum ajuste no foco (por exemplo, aprofundar uma área específica identificada no PI4). Para novos projetos, os grupos apresentam uma breve justificativa da escolha da empresa (perfil, segmento) e do problema identificado. Os professores orientam na delimitação do escopo, garantindo que esteja adequado ao nível de desafio do PI5 e factível dentro do semestre. **Entrega:** nome da empresa escolhida (ou confirmada) por cada grupo e uma *descrição breve do problema de negócio e escopo* (1 parágrafo resumindo o que se pretende resolver/analisar) – essa descrição servirá de base para refinamento no relatório. Além disso, espera-se a *criação do repositório GitHub e quadro Trello do projeto* nesta semana, se ainda não feito, com link compartilhado aos professores. Os grupos já devem agendar a primeira reunião oficial com a empresa, se possível para a próxima semana, para alinhar expectativas e coletar informações iniciais.
- Semana 3 – Kick-off do Projeto com a Empresa & Planejamento:** Realização (ou preparação) da reunião de kick-off com a empresa parceira. Nessa reunião (idealmente durante a semana 3), o grupo se apresenta formalmente, confirma o entendimento do problema de negócio, define como será a dinâmica de comunicação e validação (frequência de reuniões, pontos de contato) e levanta requisitos e dados disponíveis. Internamente, durante as aulas, discute-se planejamento detalhado do projeto: quais dados precisarão ser coletados, quais metodologias serão empregadas, divisão de tarefas inicial entre membros, e cronograma interno alinhado às entregas parciais. Os professores podem auxiliar como *facilitadores*, indicando metodologias de gestão (Canvas do projeto, 5W2H para plano de ação, cronograma macro). **Entrega:** até o

final da semana, cada grupo deve postar um *Plano de Projeto inicial*, contendo: resumo do problema e solução proposta, principais etapas/tarefas identificadas, responsáveis, cronograma preliminar e necessidades de apoio da empresa (por ex., “aguardando base de dados X até data Y”). Esse plano pode ser um documento curto ou slides, a critério, e servirá para orientar as próximas semanas. Adicionalmente, espera-se até aqui o *registro no GitHub de um README* com as informações do projeto (empresa, problema, integrantes) e criação de cartões no Trello refletindo o plano (p. ex.: tarefas de coleta de dados, tarefas de modelagem etc.).

4. **Semana 4 – Coleta de Dados (Fonte Primária e Secundária):** Os grupos iniciam a busca e obtenção dos dados necessários. Se a empresa forneceu dados internos, esta é a semana para recebê-los (respeitando NDA, se houver) e começar a analisá-los. Paralelamente, coletam-se dados secundários externos que complementem o dataset (por ex., dados públicos do IBGE, dados climatológicos, séries econômicas, textos de redes sociais etc., dependendo do projeto). Ferramentas de ETL e scripts Python podem ser empregados para automação da coleta (p.ex., usando APIs ou web scraping, se pertinente). Durante as aulas, discute-se qualidade dos dados e possíveis necessidades de limpeza ou transformação. **Entrega:** não há uma entrega formal nesta semana, mas os grupos devem registrar em relatório/diário de bordo quais fontes de dados já foram obtidas e qualquer dificuldade encontrada (por ex., necessidade de permissão adicional da empresa, ou problemas de formato). É recomendável já produzir um inventário dos dados coletados (documento listando cada conjunto de dados, origem, tamanho, conteúdo) e compartilhar com o professor para validação de que são adequados. Se algum dado crítico estiver faltando, essa é a hora de acionar a empresa ou buscar fontes alternativas.
5. **Semana 5 – Tratamento e Pré-processamento de Dados:** Com os dados brutos em mãos, inicia-se o processo de limpeza e preparação. As atividades incluem: tratar valores faltantes ou inválidos, normalizar/escalar variáveis numéricas quando necessário, codificar variáveis categóricas, integrar diferentes fontes (join/merge de tabelas), deduplicar registros, e outras etapas de data wrangling. Se houver dados textuais, aplicar técnicas de pré-processamento (remoção de ruído, tokenização, etc.). Os grupos devem aproveitar para criar notebooks de EDA inicial enquanto tratam os dados, para já ir conhecendo as principais características. Durante as orientações, os professores podem revisar parte do código de preparação de dados para garantir boas práticas (ex.: ver se estão documentando cada passo, versionando arquivos no Git, etc.). **Entrega:** uma *amostra dos dados tratados* ou descrição do pipeline de tratamento pode ser solicitada informalmente pelo professor nesta semana para feedback rápido. Em termos de artefatos, os grupos devem atualizar o repositório com os scripts/notebooks de pré-processamento e talvez gerar um primeiro conjunto de gráficos simples (histogramas, outliers) como evidência de que o tratamento foi eficaz. Essa semana é essencial para pavimentar a modelagem; portanto, o critério de sucesso é encerrar a semana com um *dataset* limpo e pronto para análise exploratória e treinamento dos modelos.
6. **Semana 6 – Análise Exploratória de Dados (EDA):** Dedicar-se esta semana a uma análise exploratória profunda do conjunto de dados já preparado. Os alunos devem investigar distribuições das variáveis, correlações, padrões temporais, segmentações etc., utilizando técnicas estatísticas e visualizações gráficas. O objetivo é extrair insights iniciais e guiar as decisões de modelagem. Por exemplo, identificar quais variáveis parecem mais influenciar o

fenômeno estudado, se há necessidade de criar variáveis derivadas (features engineering) ou de filtrar outliers que poderiam distorcer o modelo. Também é o momento de calcular estatísticas-descritivas e, se o problema for preditivo, começar a pensar no split de dados (treino/teste). **Entrega:** preparar um relatório/tabela de insights da EDA para discussão em sala ou com a empresa. Embora não seja uma entrega formal avaliada ainda, muitos grupos optam por apresentar nessa semana um *resumo da análise exploratória* para os professores e colegas, seja em formato de slides ou relatório preliminar, para receber comentários. Essa devolutiva ajuda a verificar se o entendimento dos dados está correto e se o enfoque planejado para a modelagem faz sentido. Lembre-se também de compartilhar os principais achados com a empresa – por exemplo, “*descobrimos que a região Norte tem um padrão de vendas bem distinto do Sul*”, etc. Esse feedback pode gerar um ajuste de escopo ou novos insights fornecidos pelo pessoal da empresa. Atualize o repositório GitHub com qualquer notebook de EDA produzido e considere salvar gráficos importantes para possivelmente incluir no relatório final.

7. **Semana 7 – Definição da Estratégia de Modelagem & Features:** Após compreender os dados, esta semana foca em planejar a modelagem preditiva. Os grupos devem definir quais serão as variáveis-alvo (se ainda não estiver claro), quais algoritmos de Machine Learning pretendem tentar inicialmente e quais métricas de sucesso serão usadas para avaliar os modelos, alinhando isso aos objetivos de negócio (por exemplo, “precisamos de pelo menos 85% de acurácia para que a empresa considere viável usar o modelo X”). Além disso, decide-se sobre a seleção/criação de features: com base na EDA, escolher quais atributos entrarão nos modelos, criar variáveis derivadas se útil (por exemplo, agregar valores mensais, criar categorias a partir de valores contínuos, etc.), e possivelmente reduzir dimensionalidade se houver muitas variáveis (via PCA ou seleção por importância). Essa etapa de definição é importante para dar foco: em vez de sair testando tudo sem critério, ter um plano de experimentos. **Entrega:** documento interno ou uma apresentação curta descrevendo a estratégia de modelagem – listando os modelos que serão testados, as features consideradas e o plano de validação (k-fold, porcentagem de split etc.). Envie isso ao professor para validação. Alguns grupos também já iniciam nesta semana a codificação dos primeiros modelos (*baseline* simples) para ganhar tempo, mas o grosso do treinamento acontece a partir da próxima semana. Se houver quaisquer pendências de dados ou ajustes identificados de última hora, tratar imediatamente.
8. **Semana 8 – Desenvolvimento dos Modelos Preditivos (Iteração 1):** Mãos à obra na implementação dos modelos de Machine Learning. Nesta semana, foque em desenvolver e treinar as primeiras versões dos modelos definidos: por exemplo, treinar uma regressão linear e um Random Forest (para um problema regressivo) ou uma árvore de decisão e um modelo de rede neural (para um problema classificatório), etc. Use conjuntos de validação para comparar performances e comece a documentar os resultados. Se a base de dados for grande, talvez seja necessário amostrar dados ou usar recursos computacionais adicionais (laboratórios de informática da Fatec, Google Colab, etc.). Documente também o tempo de treinamento, complexidade do modelo, e eventuais dificuldades (como necessidade de balancear classes, problemas de convergência, etc.). **Entrega:** até o fim da semana, cada grupo deve ter *pelo menos um modelo funcional treinado e avaliado preliminarmente*. Recomenda-se produzir um *checkpoint* de resultados – por exemplo, uma tabela comparativa de métricas dos modelos

testados até agora – para compartilhar com o professor/orientador e com a empresa. Essa pode ser considerada uma entrega técnica parcial: alguns grupos fazem um pequeno workshop com a empresa apresentando esses resultados iniciais, perguntando se fazem sentido e coletando sugestões (validação intermediária). Lembre-se de salvar versões do modelo e do código no GitHub (evitando riscos de perda) e começar a escrever os trechos do relatório referentes à metodologia de modelagem e às métricas, enquanto está tudo fresco na memória.

9. **Semana 9 – Iteração e Melhoria dos Modelos (Iteração 2):** Com base nos resultados iniciais, realiza-se uma segunda iteração de modelagem para melhoria. Isso pode envolver: tunar hiperparâmetros dos modelos mais promissores (via grid search, random search ou métodos automatizados); testar modelos adicionais ou variantes (por ex., verificar se adicionar uma rede neural profunda melhora significativamente ou não); e ajustar features conforme necessário (talvez removendo as que se mostraram irrelevantes ou adicionando alguma transformação não tentada antes). É provável que nesta altura um modelo “favorito” já emerja – concentre esforços nele para deixá-lo robusto. Continue avaliando com rigor, inclusive utilizando dados de teste somente ao final para ter avaliação realista. **Entrega:** ao final da semana 9 (que marca aproximadamente a metade do semestre), espera-se que cada grupo disponha de um modelo preditivo principal escolhido, com desempenho validado e pronto para ser apresentado. O grupo realizará a apresentação intermediária nessa fase: os grupos apresentam em sala de aula a primeira versão consolidada do projeto, mostrando o problema, a abordagem, o modelo escolhido e seus resultados até aqui. Isso serve como um “pilot” da apresentação final e rende feedbacks valiosos dos professores e colegas. Se essa apresentação intermediária ocorrer, encare-a como uma oportunidade de testar a narrativa e identificar pontos fracos a corrigir. Independentemente disso, compartilhe com a empresa parceira os resultados atingidos: se, por exemplo, *o modelo de churn está atingindo 90% de acerto, veja se isso é bem recebido e se há considerações* (às vezes a empresa pode dizer “preferimos privilegiar menos falsos positivos mesmo que diminua a acurácia”, etc.). Registre por escrito qualquer feedback importante.
10. **Semana 10 – Planejamento de Deploy e Protótipo:** Com o modelo definido, volta-se a atenção para o plano de implantação. Nesta semana, os grupos elaboram em detalhe a estratégia de deploy: escolhem o ambiente (nuvem ou on-premises), definem quais ferramentas serão usadas (Docker, serviços específicos), elaboram diagramas e começam a escrever o plano de deploy que irá no relatório e no guia técnico. Se possível, desenvolva um protótipo do ambiente: por exemplo, criar um pequeno serviço Flask local que carrega o modelo e fornece uma previsão via API, ou um esboço de container Docker com o modelo dentro. Não é obrigatório implementar completamente, mas qualquer avanço prático aqui ajuda a validar o plano (por ex., testar se o modelo consegue rodar em um container com 1GB de RAM, etc.). Além disso, os grupos devem consolidar todos os códigos e documentações parciais em um pacote coeso. **Entrega:** preparação de um *documento de arquitetura de deploy* (pode ser um diagrama + texto explicativo) para revisão do professor e, se possível, validação pela equipe de TI da empresa parceira. Muitos grupos encaminham esse plano para alguém da empresa opinar (“Isso faria sentido na infraestrutura de vocês?”). Em paralelo, os alunos devem iniciar a escrita do Guia de Uso da solução, pois grande parte do conteúdo estará relacionada à implantação e uso do modelo – comece a redigir instalação, dependências, exemplos de execução. Essa semana é um bom momento também para revisar os requisitos do relatório técnico e se certificar de que nenhuma parte ficou esquecida (por exemplo, garantir que já tem todos os dados necessários

para a análise de negócio, como uma atualização de indicadores da empresa se for relevante ao contexto).

- 11. Semana 11 – Elaboração do Relatório e do Guia Técnico (Versão Preliminar):** Aqui entramos na fase de documentação final. Os grupos devem dedicar a semana 11 a escrever intensivamente o relatório técnico e finalizar o guia de uso, integrando todos os elementos produzidos. Se possível, já produzir todos os gráficos, tabelas e anexos necessários e inseri-los no relatório conforme normas ABNT. É hora de garantir que o texto tenha fluidez e consistência – por exemplo, verificar se a introdução apresenta claramente o problema e se os resultados apresentados respondem à pergunta proposta, ou se todos os termos técnicos foram explicados para um leitor leigo. Os professores podem organizar uma revisão por pares: trocar relatórios preliminares entre grupos para leitura crítica, o que ajuda a pegar pontos de melhoria (às vezes um colega nota que faltou explicar certa sigla, ou que um gráfico não está legível). **Entrega:** no final desta semana, espera-se a entrega preliminar do relatório escrito e do guia de uso para feedback. Essa entrega geralmente é combinada informalmente – os alunos enviam por e-mail ou disponibilizam no GitHub uma versão “quase pronta” do relatório para que os orientadores leiam e façam comentários. Junto a isso, podem entregar também uma versão inicial dos *slides da apresentação*. Essa etapa de feedback prévio é fundamental: os professores revisarão formatação ABNT, coerência de conteúdo, seções faltantes, e retornarão na semana seguinte com apontamentos de ajuste. Lembre-se: uma boa preparação aqui evita correções de última hora na véspera da entrega final. *(Nota: A pontualidade nessa entrega preliminar pode até ser um critério avaliativo indireto, pois demonstra organização.)*
- 12. Semana 12 – Feedback e Ajustes Intermediários:** Na semana 12, os grupos recebem o feedback dos professores sobre o relatório e materiais entregues na semana anterior. Haverá reuniões de orientação para discutir as melhorias sugeridas: correções na formatação ABNT (ex.: alinhamento de capa, citações fora do padrão), complementações necessárias (ex.: “faltou explicar a equação de tal métrica” ou “seria bom incluir um gráfico comparativo dos modelos”), aperfeiçoamento da escrita (coerência, retirar informalidades) e adequação do guia de uso (ex.: “faltou informar versão do Python usado”). Esta semana também pode ser utilizada para uma *apresentação simulada* – os grupos ensaiam a apresentação final diante dos colegas e professores, cronometrando o tempo e recebendo críticas construtivas quanto à clareza e didática. Se possível, envolva alguém da empresa nesse ensaio ou como público-alvo para garantir que a linguagem está acessível a todos. **Entrega:** não há uma nova entrega formal a ser enviada nesta semana (já que a preliminar foi na anterior), mas é imprescindível incorporar todas as sugestões de melhoria apontadas. Atualize o relatório e slides conforme indicado, e refine o produto técnico se necessário (por exemplo, corrigir um bug no código descoberto durante testes de uso). Garanta também que todos os anexos exigidos (atas, validações da empresa etc.) estejam prontos ou em vias de ficar – se falta alguma assinatura da empresa, corra atrás agora.
- 13. Semana 13 – Versão Final do Relatório e Preparativos Finais:** Chegou a reta final para a entrega escrita. Até o meio da semana 13, cada grupo deve finalizar 100% do relatório técnico e do guia de uso, revisando minuciosamente ortografia, formatação e completude. Conferir se todas as seções estão presentes, se todas as referências citadas no texto estão listadas nas Referências e vice-versa, se os anexos estão numerados e referenciados no corpo do texto, etc. Atenção também à encadernação ou submissão digital conforme instruído – a definição deve

ocorrer em reunião de colegiado/NDE. Paralelamente, os slides da apresentação devem ser finalizados, incorporando eventuais mudanças do relatório final. **Entrega:** normalmente, até o final da semana 13 (ex.: sexta-feira) é o prazo para entrega oficial do Relatório Técnico final e do Guia de Uso aos professores. Essa é a versão que será avaliada pela banca, então trate-a como definitiva. Envie nos formatos solicitados (geralmente PDF para o relatório; o guia pode ser PDF também; slides podem ser enviados em PDF ou no formato de apresentação se requerido). Certifique-se de que o repositório GitHub esteja atualizado e organizado – muitas vezes os avaliadores podem consultar o repositório durante a banca, então garanta que ele reflète fielmente o conteúdo final (código atualizado, README descritivo etc.). Com a documentação entregue, foque nos preparativos da apresentação: divisão de falas, tempo de cada membro, esclarecimento de quem responde sobre qual parte nas perguntas etc.

**14. Semana 14 – Preparação Final da Apresentação (com orientação do professor):** Nesta semana, o foco será *exclusivamente a preparação final para a apresentação do Projeto Integrador V*, com acompanhamento direto do professor responsável pela disciplina. É o momento de consolidar tudo o que será apresentado, eliminando dúvidas e ajustando os últimos detalhes técnicos e de comunicação. Entrega: Revisão e refinamento do conteúdo da apresentação (roteiro, narrativa, tempo e objetivos). Ajustes finais nos materiais de apoio (slides, roteiro de fala, demonstração do produto, prints, dashboards, notebooks). Validação técnica do que será demonstrado (testes, dados de exemplo, ambiente configurado, plano B para falhas). **Contato com a empresa parceira**, alinhando expectativas, confirmando o recorte do projeto e definindo: o que será mostrado (escopo final), como será apresentado (formato e dinâmica), quais resultados e impactos devem ser priorizados. Ensaios orientados (com feedback do professor) para melhorar postura, clareza, objetividade e domínio do conteúdo. Ao final da semana, o grupo deve ter *apresentação e demonstração prontas*, com tudo revisado e consistente com a documentação/relatório do projeto.

**15. Semana 15 – Apresentação Final do PI5 e Encerramento:** Nesta semana ocorrerá a apresentação final do PI5, na qual os grupos deverão expor de forma completa e profissional a solução desenvolvida, incluindo contexto do problema, abordagem adotada, resultados obtidos e demonstração do produto/artefato final. **Dinâmica sugerida:** a) Apresentação formal do projeto (com tempo definido pela banca). b) Demonstração objetiva do produto/solução (preferencialmente com um caso de uso simples e claro). c) Arguição da banca e respostas do grupo. d) Encerramento com agradecimentos à empresa/parceiros e reflexão breve sobre próximos passos (continuidade no PI6). Após as apresentações, a banca poderá deliberar notas e, quando aplicável, fornecer feedback resumido (pontos fortes e melhorias). Recomenda-se que, ao término, o grupo organize e publique os materiais finais no GitHub, garantindo repositório limpo, completo e coerente com o que foi apresentado.

*(Observação: O cronograma acima é sugerido e pode ser ajustado conforme calendário específico ou imprevistos. O importante é não postergar etapas críticas – por exemplo, a coleta de dados deve acontecer o quanto antes, e a versão preliminar do relatório deve ser levada a sério – para evitar acúmulo de trabalho no final. Planejamento e disciplina são chaves para cumprir todas as etapas no prazo.)*

# Sistema de Avaliação

O sistema de avaliação do Projeto Integrador V totaliza **100 pontos**, distribuídos entre vários critérios, de forma a abranger tanto o processo quanto os produtos finais do projeto. A tabela de pontuação é definida na ficha de avaliação oficial e está prevista aproximadamente da seguinte maneira:

- **Participação e evidências de interação com a empresa (10 pts):** Avalia o compromisso do grupo com a extensão, medido pela qualidade e frequência dos *registros de reuniões* com a empresa parceira e pela colaboração demonstrada. Para pontuação máxima, espera-se atas bem elaboradas de diversos encontros e retorno positivo da empresa quanto ao engajamento dos alunos.
- **Entregas técnicas parciais validadas (10 pts):** Refere-se às *entregas intermediárias* realizadas e aprovadas pela empresa durante o semestre. Inclui a pontualidade e completude dessas entregas (ex.: envio de análise exploratória, apresentação de resultado preliminar, protótipo de modelo) e a obtenção de feedback/assinatura da empresa em cada marco. Grupos que seguiram a abordagem de sprints curtos com devolutivas frequentes tendem a alcançar nota alta aqui, enquanto ausência de validações ou entregas atrasadas reduzem a pontuação.
- **Relatório técnico e guia de uso (20 pts):** Avalia a qualidade do relatório escrito e do manual técnico entregues. Os avaliadores verificarão a estrutura, formatação e conformidade com normas ABNT, a clareza da redação, a coerência e profundidade da análise apresentada, além da organização e utilidade do guia de uso. Itens como revisão bibliográfica pertinente, discussão dos resultados, qualidade de gráficos/tabelas, e ausência de erros ortográficos/pontuação influenciam essa nota. Um relatório exemplar é aquele completo, bem escrito e corretamente formatado, que demonstra rigor acadêmico e reflexão crítica.
- **Apresentação oral (30 pts):** Corresponde ao desempenho na banca de apresentação. Serão considerados a comunicação oral (clareza, objetividade, entrosamento da equipe, distribuição de falas), a qualidade dos slides (visual limpo, conteúdo sintetizado, uso adequado de figuras), a gestão do tempo e a capacidade de responder perguntas mostrando domínio do projeto. Também conta a postura profissional – segurança, respeito e trabalho em equipe durante a apresentação. Mesmo valendo menos pontos que outras partes, uma apresentação ruim pode prejudicar a compreensão do projeto pelos avaliadores, então deve-se buscar excelência também neste quesito.
- **Produto funcional de ciência de dados (30 pts):** Mede a qualidade e efetividade do produto técnico entregue – código, modelo e demais componentes da solução. Será avaliado se o pipeline de análise/modelagem foi bem implementado, se os resultados (modelo preditivo, dashboards) atendem aos objetivos propostos e se a solução é *reprodutível* e bem documentada. Critérios específicos incluem: uso adequado de técnicas de ML/PLN (aspecto técnico), inovação ou complexidade da solução, e aderência às necessidades da empresa (aspecto prático). A banca pode inspecionar o repositório GitHub e executar parcialmente o código para verificar funcionamento. Adicionalmente, a presença de componentes extras como um protótipo de deploy (ex.: container, API demonstrativa) conta positivamente. Em

resumo, esse critério reflete o “valor agregado” pelo projeto em termos de resultado de ciência de dados.

A soma dos itens acima perfaz 100 pontos. Cada integrante do grupo recebe, em princípio, a mesma nota do grupo, a menos que haja evidências documentadas de contribuição muito discrepante (em casos raros, professores podem ajustar notas individuais se um membro não participou adequadamente, usando como base o histórico de commits, depoimentos dos colegas e autoavaliações). O Projeto Integrador V terá um *peso significativo na nota das disciplinas integradas*, compondo tipicamente **20% da média final** de cada uma delas. Ou seja, o desempenho no PI5 impacta diretamente as notas de Aprendizado de Máquina, PLN, Análise Preditiva e Infraestrutura, reforçando a importância de se dedicar proporcionalmente a esta atividade.

A avaliação do PI5 será conduzida de forma transparente e objetiva, com base em critérios detalhados na ficha avaliativa. Recomenda-se que os grupos conheçam antecipadamente esses critérios (geralmente fornecidos no início do semestre) e até os usem como *checklist* antes da entrega – por exemplo, checando item por item se atenderam a todos os pontos esperados. Normalmente, ao final da banca, os avaliadores preenchem a ficha atribuindo notas por critério e justificativas sucintas; em seguida, essas fichas podem ser disponibilizadas aos grupos para conhecimento das forças e fraquezas de seu projeto.

**Dica:** *Para alcançar desempenho máximo, os grupos devem focar na qualidade em todas as frentes: relatório em conformidade com normas ABNT, código bem escrito e documentado no GitHub, modelo preditivo com desempenho consistente, slides claros e visuais, e envolvimento ativo com a empresa evidenciado por reuniões e entregas. Entre as boas práticas estão: definir um problema de negócio claro e relevante, triangular fontes de informação (dados internos, externos, literatura) para embasar a solução, documentar todas as referências utilizadas, manter uma linha lógica que conecte análise de dados e estratégia (ex.: resultados do modelo levando a recomendações de negócio), incluir ao menos um elemento técnico inovador aprendido nas disciplinas (por exemplo, se possível aplicar um modelo de NLP ou uma técnica avançada de ML para mostrar domínio), testar a solução exaustivamente antes da entrega (garantir que não há falhas de execução), e apresentar-se de forma coesa e confiante na banca. O uso efetivo das ferramentas (GitHub, Trello) e o acompanhamento próximo do orientador durante todo o processo também são fundamentais – grupos organizados raramente são pegos de surpresa ou deixam requisitos de fora.*

## ANEXO A

# Ficha de Avaliação – Apresentação Final

### Projeto Integrador V – Projetos em Ciência de Dados I

**Curso:** CST em Ciência de Dados para Negócios – Fatec Votorantim

**Semestre:** 5º semestre

**Grupo:** \_\_\_\_\_

**Empresa Parceira:** \_\_\_\_\_

**Data:** \_\_\_\_ / \_\_\_\_ / \_\_\_\_

**Avaliadores:** \_\_\_\_\_

#### 1. Comunicação e Estrutura da Apresentação (10 pontos)

Critério	Descrição	Pontuação
Clareza e organização	Apresentação segue sequência lógica (problema → metodologia → resultados → deploy).	0 – 2
Objetividade	Conteúdo apresentado de forma clara, sem excesso ou lacunas relevantes.	0 – 2
Gestão do tempo	Respeito ao tempo estabelecido, com boa distribuição entre os membros.	0 – 2
Qualidade visual dos slides	Slides legíveis, organizados e adequados ao padrão institucional.	0 – 2
Coerência narrativa	A apresentação conta a “história do projeto” de forma compreensível.	0 – 2

**Subtotal:** \_\_\_\_\_ / 10

#### 2. Domínio Técnico e Metodológico (10 pontos)

Critério	Descrição	Pontuação
Compreensão do problema de negócio	Demonstra entendimento claro da demanda da empresa.	0 – 2
Explicação da metodologia	Pipeline de ciência de dados explicado corretamente.	0 – 2
Modelagem preditiva	Justificativa adequada para escolha dos modelos utilizados.	0 – 2
Interpretação dos resultados	Métricas e resultados explicados de forma correta e crítica.	0 – 2
Integração interdisciplinar	Evidência de integração das disciplinas do semestre.	0 – 2

**Subtotal:** \_\_\_\_\_ / 10

#### 3. Demonstração da Solução Desenvolvida (5 pontos)

Critério	Descrição	Pontuação
Funcionamento do produto	Demonstração clara do modelo/dashboard/protótipo.	0 – 2
Aplicabilidade prática	Evidência de utilidade para a empresa parceira.	0 – 2
Organização técnica	Demonstração estruturada e compreensível.	0 – 1

**Subtotal:** \_\_\_\_\_ / 5

#### 4. Planejamento de Deploy e Visão Profissional (3 pontos)

<b>Critério</b>	<b>Descrição</b>	<b>Pontuação</b>
Clareza do plano de implantação	Explicação da arquitetura e uso futuro da solução.	0 – 2
Viabilidade prática	Proposta alinhada à realidade da empresa.	0 – 1

**Subtotal:** \_\_\_\_\_ / 3

---

#### 5. Postura Profissional e Trabalho em Equipe (2 pontos)

<b>Critério</b>	<b>Descrição</b>	<b>Pontuação</b>
Participação equilibrada	Distribuição adequada das falas entre os integrantes.	0 – 1
Postura e comunicação	Segurança, respeito e comportamento profissional.	0 – 1

**Subtotal:** \_\_\_\_\_ / 2

---

#### Pontuação Total da Apresentação

**Nota Final:** \_\_\_\_\_ / 30 pontos

---

#### Comentários da Banca Avaliadora

---

---

---

---

# Planejamento das Atividades de Extensão Universitária

## Projeto Integrador V – Projetos em Ciência de Dados I

---

### 1. Identificação da Atividade Extensionista

**Título da Atividade:**

**Propostas de Soluções de Ciência de Dados com Modelagem e Deploy para Empresas da Comunidade Regional**

**Temática:**

Extensão tecnológica • Ciência de Dados Aplicada • Modelagem Preditiva • Estratégias de Deploy • Soluções orientadas por dados.

**Componente Curricular:**

Projeto Integrador V — Projetos em Ciência de Dados I

**Curso:**

CST em Ciência de Dados para Negócios — Fatec Votorantim

**Carga Horária:**

80 aulas (66,6 horas) — correspondente a 100% da carga horária da disciplina.

---

### 2. Caracterização da Extensão

A atividade extensionista do Projeto Integrador V consiste no desenvolvimento de propostas aplicadas de soluções em Ciência de Dados voltadas a demandas reais de empresas da comunidade regional. A proposta integra ensino, prática profissional e interação social, promovendo a aplicação dos conhecimentos acadêmicos em contextos organizacionais concretos.

Os estudantes atuarão diretamente com empresas parceiras, realizando diagnóstico do problema, definição da abordagem analítica e elaboração de uma solução completa baseada em dados, contemplando modelagem preditiva e planejamento de implantação (deploy). O foco central está na construção de uma proposta tecnicamente fundamentada, viável e aplicável, apresentada tanto à empresa quanto à comunidade acadêmica.

A atividade caracteriza-se como extensão universitária por promover troca efetiva de conhecimentos entre instituição e sociedade, contribuindo simultaneamente para a formação profissional dos estudantes e para a disseminação do uso estratégico da ciência de dados no ambiente empresarial regional.

---

### 3. Objetivos da Atividade Extensionista

- Fortalecer a integração entre a Fatec Votorantim e empresas da Região Metropolitana de Sorocaba;
- Promover a aplicação prática da Ciência de Dados em problemas empresariais reais;
- Planejar soluções completas orientadas por dados, incluindo pipelines analíticos, modelos preditivos e estratégias de deploy;
- Desenvolver competências técnicas, analíticas e comunicacionais dos estudantes;

- Oferecer às empresas participantes propostas estruturadas de alto valor agregado, passíveis de futura implementação.
- 

#### **4. Público-Alvo**

- Empresas locais e organizações da Região Metropolitana de Sorocaba;
  - Comunidade acadêmica da Fatec Votorantim;
  - Profissionais das áreas de tecnologia, dados e negócios interessados nas soluções desenvolvidas.
- 

#### **5. Etapas e Ações de Execução**

A atividade extensionista será desenvolvida ao longo do semestre, seguindo as etapas abaixo:

1. Seleção das empresas parceiras e organização dos grupos de estudantes;
  2. Realização de reuniões de diagnóstico para compreensão do negócio e identificação do problema real;
  3. Definição da pergunta analítica e levantamento dos dados necessários;
  4. Planejamento do pipeline de Ciência de Dados, incluindo:
    - a. coleta e tratamento de dados;
    - b. engenharia de atributos;
    - c. seleção e definição de modelos preditivos;
    - d. definição das métricas de avaliação;
    - e. elaboração da estratégia de deploy (ambiente local, servidor ou nuvem);
  5. Elaboração do relatório técnico e do guia da solução proposta;
  6. Produção de vídeo pitch apresentando a solução desenvolvida;
  7. Apresentação final presencial à empresa parceira e à comunidade acadêmica, com devolutiva técnica.
- 

#### **6. Entregas Previstas**

- Relatório técnico completo contendo:
    - contextualização do problema;
    - pipeline proposto e justificativas técnicas;
    - estratégia de validação e métricas;
    - plano de deploy da solução;
  - Guia prático de implementação;
  - Vídeo pitch explicativo da proposta;
  - Slides de apresentação final;
  - Registros documentais das interações com a empresa parceira.
- 

#### **7. Instrumentos e Procedimentos de Avaliação**

A avaliação das atividades extensionistas ocorrerá de forma contínua, considerando:

- rubrica avaliativa do relatório técnico;
  - rubrica de avaliação do vídeo pitch;
  - avaliação da apresentação pública final (clareza, aplicabilidade e domínio técnico);
  - observação da participação dos estudantes nas reuniões e interações com a empresa;
  - feedback formal da empresa parceira, coletado por meio de formulário estruturado.
-

## **8. Formas de Evidência da Extensão**

As atividades extensionistas deverão ser comprovadas por meio dos seguintes registros:

- relatórios técnicos, vídeos e apresentações produzidos;
- atas ou registros das reuniões com empresas parceiras;
- capturas de tela e registros dos sistemas e ferramentas utilizadas;
- guias técnicos e documentação gerada pelos grupos;
- formulários de feedback das empresas participantes.

Essas evidências deverão demonstrar a interação efetiva com o parceiro externo e a aplicação prática do conhecimento acadêmico em contexto real.

---

## **9. Resultados Esperados**

Espera-se que a atividade extensionista:

- contribua para a disseminação da cultura de tomada de decisão orientada por dados na comunidade regional;
- fortaleça o vínculo institucional entre a Fatec Votorantim e o setor produtivo;
- desenvolva nos estudantes competências profissionais alinhadas às demandas do mercado;
- gere propostas de soluções tecnológicas com potencial de implementação futura pelas empresas parceiras.

# Resumo Executivo — Projeto Integrador V (PI5) • CDN

**Propósito:** O PI5 (“Projetos em Ciência de Dados I”) integra as disciplinas do 5º semestre do curso (Aprendizado de Máquina Aplicado à CDN I, PLN Aplicada à CDN, Análise Preditiva para CDN e Infraestrutura para CDN e Big Data I) para desafiar os estudantes a desenvolverem uma solução completa em Ciência de Dados para um problema real de negócio. Isso inclui a construção de modelos preditivos robustos e o planejamento de deploy desses modelos em ambiente de produção, preferencialmente em parceria com empresas reais. O projeto busca consolidar a visão de negócio com habilidades técnicas avançadas, estimulando pensamento analítico, capacidade de modelagem e planejamento de infraestrutura, sempre alinhados às necessidades da empresa parceira e aos padrões profissionais e acadêmicos. Trata-se de uma atividade de extensão que aproxima a teoria da prática, conectando academia e setor produtivo, em conformidade com as diretrizes nacionais de educação profissional (Resolução CNE/CP 1/2021, Art. 30, incisos I, VIII, IX, X). Em suma, o PI5 leva os alunos a *proporem soluções de Ciência de Dados end-to-end, do entendimento do problema à preparação para implantação, desenvolvendo competências técnicas, estratégicas e de comunicação profissional.*

### Diretrizes-chave:

- **Empresa parceira real (obrigatório):** Cada grupo deve trabalhar com uma empresa real, preferencialmente que atue no entorno da Fatec Votorantim (RMS – Região Metropolitana de Sorocaba), sem repetição de empresa entre grupos. Recomenda-se fortemente manter a mesma empresa parceira do PI4, dando continuidade ao projeto anterior para aprofundar a solução. Caso opte-se por um novo projeto, a empresa e o problema escolhidos devem atender aos critérios de relevância e disponibilidade de dados, envolvendo um desafio real de negócio (validação pelos professores na semana inicial).
- **Grupos:** Compostos por até 6 integrantes, preferencialmente os mesmos do PI4 para preservar a sinergia e experiência acumulada. A colaboração em equipe é essencial, com divisão de tarefas equilibrada conforme as habilidades de cada membro. Mudanças de integrantes ou de grupo devem ser excepcionais e aprovadas pelo(a) professor(a) responsável pelo Projeto Integrador.
- **Portfólio e gestão do projeto:** Toda a documentação deve ser centralizada em um repositório *GitHub* do grupo (controle de versão), e o planejamento das tarefas deve ser conduzido com uso de sistema gerenciador de projetos, como por exemplo o *Trello* (preferencialmente utilizando a metodologia *kanban*) ou ferramenta similar. Espera-se registro contínuo do progresso (commits frequentes no *GitHub*, issues/pull requests etc.) e uso de quadros ágeis para acompanhamento das atividades e sprints. Essas ferramentas permitem transparência e organização, sendo parte dos critérios de avaliação do projeto.
- **Dados & Ferramentas:** O projeto exigirá uso intensivo de técnicas e ferramentas de ciência de dados avançadas. Os grupos deverão empregar *bibliotecas de Python para machine learning* (por exemplo, *Scikit-learn*, *statsmodels* etc.), ferramentas de *Processamento de Linguagem Natural (NLP)* se lidarem com dados textuais (*NLTK*, *SpaCy*, *LLMs* etc.), além de frameworks de análise de dados como *Pandas/NumPy* para preparação de dados. Conceitos de *infraestrutura e Big Data* aprendidos (como uso de bancos de dados, *cloud computing*, *containers* ou *pipelines de dados*) devem ser incorporados no planejamento de deploy. A coleta de dados deve abranger fontes primárias e secundárias confiáveis, garantindo volume e qualidade suficientes para a modelagem preditiva.
- **Conformidade acadêmica:** Todos os artefatos escritos do projeto devem seguir rigorosamente as normas institucionais e da ABNT. O relatório técnico final deve usar o template oficial da Fatec (capa, fontes, margens) e incluir citações e referências no formato ABNT. Da mesma forma, a *apresentação oral* deve utilizar o modelo de slides institucional, prezando pela clareza visual e aderência ao tempo estipulado. A escrita deve ser formal e objetiva, evitando coloquialismos, e todas as fontes de pesquisa ou dados utilizados devem ser devidamente referenciadas.

- **Ética e profissionalismo:** Os alunos devem conduzir o projeto de forma ética, respeitando a confidencialidade e privacidade dos dados da empresa. Se necessário, assinar termos de **NDA (Non-Disclosure Agreement)** com a empresa para uso de dados sensíveis. É crucial seguir os princípios da LGPD (Lei Geral de Proteção de Dados) ao manusear dados pessoais. Em todas as interações com a empresa e demais stakeholders, espera-se postura profissional, pontualidade em reuniões e cumprimento dos acordos. Todas as decisões tomadas no projeto devem considerar aspectos éticos (ex.: evitar vieses discriminatórios nos modelos) e a responsabilidade socioambiental. Além disso, deve-se evitar plágio, produzindo conteúdo original no relatório e citando adequadamente quaisquer referências teóricas ou técnicas utilizadas.

Em uma frase:

***o PI5 exige que o grupo resolva um problema real de negócio através de uma solução preditiva de ciência de dados e planeje sua implantação, demonstrando excelência técnica, visão prática e rigor na comunicação – tudo isso evidenciado por um processo bem documentado e alinhado aos padrões profissionais e acadêmicos.***