

formação em dados



ds

RELATÓRIO

PARCIAL

Entrega 3

Dex: DS7 - Grupo 19

RID: 89827 - Paulo Rafael do Nascimento Lima

RID: 91455 - Rafael Pereira do Nascimento

RID: 95679 - Ricardo Kendy Nakazone

RID: 82272 - Renato Oliveira

RID: 86790 - Tibério Rocha e Cruz

RID: 97581 - Valdir Nascimento dos Santos

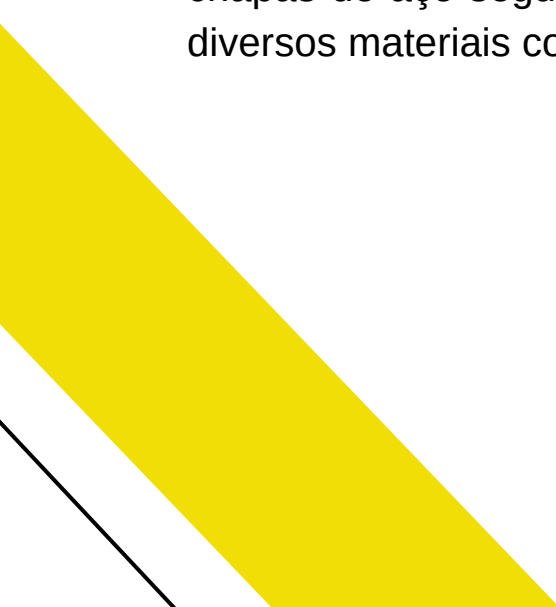
BUSINESS UNDERSTANDING

1.1 OBJETIVO DO NEGÓCIO

A empresa Tenaris é uma das maiores produtoras de tubos e aço do mundo, oferecendo uma ampla gama de produtos para atender diversas necessidades de seus clientes, desde tubos para perfuração de poços de petróleo até tubos para aplicações estruturais. Além de serviços de valor agregado, como revestimentos de tubos e serviços de inspeção.

A concorrência da Tenaris é forte, mas a empresa tem uma posição competitiva sólida no mercado global de tubos e serviços para a indústria de energia, graças à sua inovação contínua, investimentos em tecnologia, ampla presença global e compromisso com a qualidade e eficiência operacional. Além disso, a empresa apresenta soluções criativas e sustentáveis, e busca continuamente a excelência em suas operações e no relacionamento com as partes interessadas.

A Tenaris Confab é uma unidade localizada no Brasil e fabrica tubos soldados com costura. Este processo industrial inclui moldagem de chapas de aço seguidas de soldagem. Para tal processo é utilizado diversos materiais como aço, arame de solda e outros insumos.

A thick yellow diagonal bar runs from the bottom-left corner towards the center of the page.

BUSINESS UNDERSTANDING

A quantidade de insumo utilizado no processo está relacionado diretamente no valor dos Standard (STD), custos produtivos de uma oferta ao Comercial, e é definida por faixas de acordo com as características do produto.

O objetivo do projeto é desenvolver um modelo preditivo para calcular o fator de consumo para os principais insumos, utilizando dados disponibilizados pela própria empresa, a fim de obter valores mais próximos da realidade e reduzir os desvios produtivos. A análise contará com diversas etapas, incluindo exploração de dados, exploração analítica, limpeza de dados, construção de modelos, testes.

Atualmente a empresa possui um modelo que prediz um valor distante do real, como métrica de sucesso, o grupo deverá criar um modelo que consiga prever valores entre o real e a atual predição utilizada na empresa Tenaris.

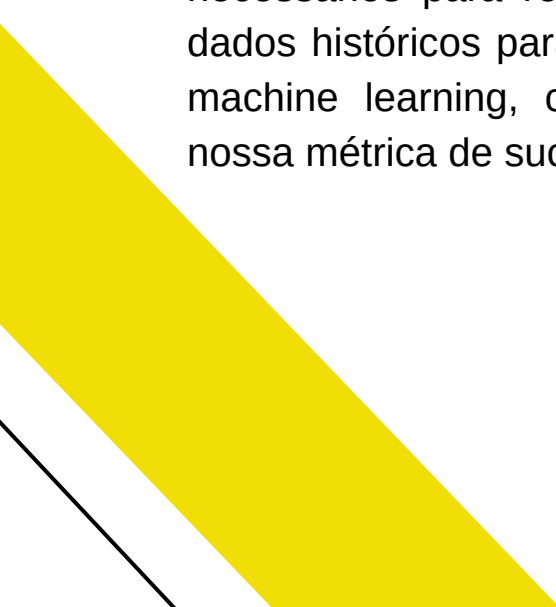
BUSINESS UNDERSTANDING

1.2 AVALIANDO A SITUAÇÃO

Para esclarecer a situação do negócio, realizou-se uma entrevista com o patrocinador do projeto. Durante a entrevista, ele explicou o modelo de negócio da empresa, a estrutura de vendas e as áreas que precisam ser melhoradas. Com essas informações, obtivemos conhecimento sobre a metodologia de cobrança da empresa e as técnicas utilizadas para prever os custos do projeto. Para ilustrar o modelo de negócio, a empresa fabrica produtos sob demanda e, portanto, precisa calcular o custo do item vendido antes de produzi-lo, uma vez que o preço final é acordado durante a venda. Nesse sentido, é essencial fazer uma estimativa de custos precisa e próxima da realidade futura. Durante a entrevista, também foi mencionado que a empresa possui um desvio padrão médio de 3% a 4%. Nossa métrica de sucesso consiste em reduzir esse percentual.

1.2.1 Inventário de recursos:

Realizamos um alinhamento para entender quais recursos serão necessários para realizar o projeto. Sabemos que precisamos de dados históricos para embasar nosso trabalho e utilizar técnicas de machine learning, como o modelo de regressão, para alcançar nossa métrica de sucesso.



BUSINESS UNDERSTANDING

Nesse sentido, levantamos hipóteses sobre os recursos necessários, como especialistas na área, ferramentas e equipamentos, como um banco de dados e IDEs para manipulação desses dados. A escolha da linguagem e frameworks a serem utilizados será discutida na próxima sprint.

1.2.2 Requisitos, premissas e restrições:

Durante a entrevista com o patrocinador, foram definidos os requisitos do projeto e incluído o cronograma de conclusão. Também foram discutidas questões de segurança, uma vez que teremos acesso a dados restritos da empresa. O patrocinador concordou em criar uma view para que possamos consumir os dados via API. Questões financeiras não foram mencionadas e acreditamos que não sejam relevantes para esse projeto em particular. Algumas premissas precisam ser observadas:

1. Disponibilidade de dados relevantes: O projeto depende da disponibilidade de dados históricos precisos e relevantes sobre o consumo de insumos e outras variáveis relevantes.
2. Acesso a tecnologias de mineração de dados: O projeto exige a utilização de tecnologias de mineração de dados, incluindo algoritmos de Machine Learning e ferramentas de visualização de dados.

BUSINESS UNDERSTANDING

3. Conhecimento do negócio: A equipe do projeto deve ter um bom conhecimento do negócio e do setor em que a empresa atua. Isso ajudará a entender as nuances dos dados e a identificar padrões que possam ser relevantes para o projeto.
4. Definição clara de objetivos: É importante que o objetivo do projeto seja claramente definido e acordado com as partes interessadas da empresa.
5. Compromisso da empresa: A empresa deve estar comprometida em implementar as mudanças necessárias com base nos resultados do projeto.
6. Gerenciamento de riscos: O projeto pode envolver riscos, incluindo a possibilidade de resultados imprecisos ou falhas na implementação das mudanças necessárias.


A Tenaris presa por muitas coisas que contribuem com a sociedade e com o cliente, podemos destacar algumas premissas da empresa:

7. A qualidade dos produtos é fundamental para garantir a segurança e a eficiência do sistema de tubulação.
8. A inovação é crucial para a competitividade e o sucesso da empresa no mercado global.
9. A personalização das soluções oferecidas é essencial para atender às necessidades específicas de cada cliente.

BUSINESS UNDERSTANDING

1.2.3 Riscos e contingência:

O sponsor se mostrou disponível para nos ajudar sempre que necessário e sanar quaisquer dúvidas que possam comprometer a entrega do projeto. A fim de minimizar os impactos no projeto, ele nos forneceu uma visão geral do dia a dia de trabalho e esclareceu, ponto a ponto, os objetivos do projeto. Desta forma segue o que entendemos como riscos ao projeto:

1. Dados insuficientes ou de baixa qualidade: um dos principais riscos é a falta de dados suficientes ou a baixa qualidade dos dados disponíveis.
 2. Viés nos dados: o viés nos dados pode afetar a precisão e a imparcialidade do modelo.
 3. Falta de compreensão do domínio: é importante que os especialistas de domínio participem do desenvolvimento do projeto de machine learning (ML).
 4. Falta de capacitação da equipe: Como estamos em uma caminhada de aprendizado, pode ocorrer que o pouco conhecimento impeça uma melhor qualidade no desenvolvimento do ML.
- 


BUSINESS UNDERSTANDING

5. Falha na implementação: a implementação do modelo é uma parte crítica do projeto de ML, pelos mesmos motivos elencados no item anterior.
6. Regulamentação e conformidade: a utilização de ML em empresas pode estar sujeita a regulamentações e conformidades específicas do setor, como leis de proteção de dados e regulamentações de segurança. O Sponsor inclusive relatou que poderia retirar alguns dados sensíveis mesmo com o acordo de confidencialidade entre a Tenaris e a DNC.
7. Perda de dados e documentos produzidos: Arquivos se corrompem ou simplesmente ficam indisponíveis.

Para contingenciar e mitigar estes riscos, é possível tomar as seguintes medidas:

1. Coletar e avaliar dados de qualidade: é importante garantir que os dados coletados sejam de alta qualidade e suficientes para suportar o projeto de ML.
2. Identificar e mitigar o viés nos dados: é importante identificar o viés nos dados e tomar medidas para mitigá-lo. Isso pode incluir a coleta de dados adicionais, a utilização de técnicas de amostragem aleatória e a avaliação cuidadosa dos algoritmos de ML para garantir que não incorram em viés.

BUSINESS UNDERSTANDING

3. Envolver especialistas de domínio: O Sponsor demonstrou muito domínio da situação, inclusive explicou o funcionamento da produção com destreza de engenheiro, mas se algum dado exigir um conhecimento técnico, solicitaremos do mesmo um parecer do responsável pela área deste dado.
 4. Capacitar a equipe: Estamos estudando e aprendendo, mas pode ser necessário buscar outras fontes de conhecimento, inclusive de IA disponíveis no mercado.
 5. Realizar testes de validação: é importante realizar testes de validação cuidadosos para garantir que o modelo funcione conforme o esperado.
 6. Manter a conformidade regulamentar: é importante garantir que o projeto de ML esteja em conformidade com as regulamentações e as políticas da empresa. Nas próximas reuniões, buscaremos conhecer as políticas da mesma e a regulamentação pública do setor.
 7. Preparar planos de contingência: Há risco de perda de dados e para isso, faremos backup de todas as informações produzidas e disponibilizadas pela empresa, e também, para os documentos que nós produzirmos estarão depositadas no git e em outras ferramentas colaborativas como teams e o google suite.
- 

BUSINESS UNDERSTANDING

1.2.4 Terminologia

A terminologia do projeto foi apresentada pelo sponsor com o objetivo de esclarecer os termos e conceitos que serão utilizados no projeto para os desenvolvedores.

Alguns termos importantes:

- Standard - STD
- Custo STD - Qtde de Material (histórico) x Preço do Material
- Fator de Consumo (FC) - Qtde de Material p/ produzir 1 kg de tubo
- Confab - É a unidade da Tenaris no Brasil que fabrica tubos soldados

1.2.5 Custos e benefícios:

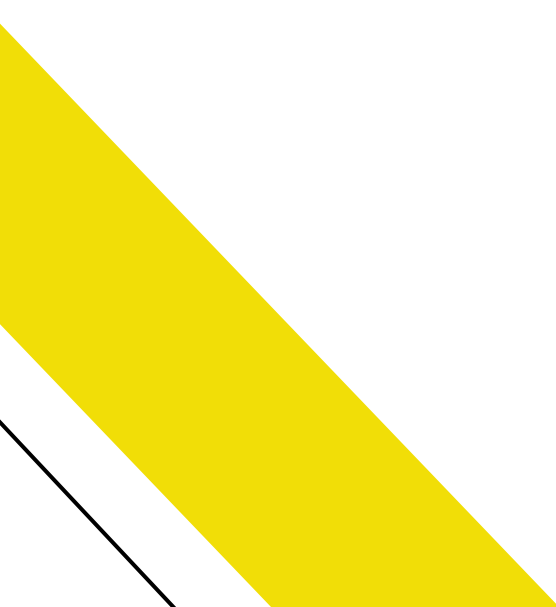
Durante a entrevista com o sponsor, não foi abordada a questão dos custos do projeto, mas a equipe estimou a quantidade de horas necessárias para a sua realização. Além disso, também foi calculado o custo por hora dos especialistas envolvidos no projeto, a fim de ilustrar um custo geral. Com relação aos benefícios para o patrocinador do projeto, a equipe espera reduzir o desvio padrão médio do fator de consumo do cliente, o que resultará em uma margem mais saudável e previsível nas negociações futuras. Embora não tenha sido possível acessar informações confidenciais, como margem ou lucro da empresa, é certo que o sucesso do projeto trará frutos positivos para a Tenaris. No entanto, não foi possível calcular um ROI (retorno do investimento).

BUSINESS UNDERSTANDING

1.3 DATA MINING GOALS

A empresa adota o modelo de vendas personalizado de Taylor, que busca entender as necessidades e preferências individuais de cada cliente para adaptar a abordagem de vendas. Essa técnica pode aumentar a confiança do cliente e ajudá-lo a tomar uma decisão de compra informada, mas exige um investimento maior em treinamento e desenvolvimento de habilidades para os vendedores.

No entanto, essa abordagem personalizada resulta em um desafio para a empresa, pois ela não pode trabalhar com estoque de produtos, já que cada item é feito sob medida para cada cliente. Isso significa que a Tenaris não consegue diluir as oscilações nos custos dos produtos em materiais que possuem em estoque, o que pode afetar a margem de lucro da empresa. Para manter a qualidade e a personalização dos produtos, a empresa precisa repassar essa diferença de custo diretamente para o cliente ou aceitar uma redução nos lucros.



BUSINESS UNDERSTANDING

1.3.1 Alguns goals podem ser elencados:

1. Aumento da eficiência operacional: o uso de um modelo de ML para otimizar o consumo de insumos pode levar a uma maior eficiência operacional, reduzindo o desperdício e os custos associados. Isso pode ser medido pela redução dos custos dos insumos e pela melhoria da produção.
2. Identificar os principais insumos que têm maior impacto no desvio padrão do fator de consumo e entender as causas subjacentes.
3. Identificar padrões nos dados históricos que possam ser usados para prever o consumo futuro de insumos e, portanto, reduzir a variação no fator de consumo.
4. Analisar os dados do processo de produção e identificar pontos de melhoria para reduzir a variação no consumo de insumos.
5. Identificar qualquer correlação entre o consumo de diferentes insumos e entender como isso pode ser usado para otimizar o consumo de insumos.
6. Identificar quaisquer fatores externos (como mudanças nas condições do mercado) que possam estar contribuindo para a variação no fator de consumo e avaliar como esses fatores podem ser gerenciados.
7. Desenvolver modelos preditivos precisos para prever o consumo de insumos com antecedência e reduzir a variação no fator de consumo.

BUSINESS UNDERSTANDING

1.3.2 Problema e objetivo do negócio

Apesar de a Tenaris já estar ciente do desafio de lidar com variações de custos devido à personalização de cada produto, a empresa criou um fator de custo para minimizar os impactos dessas variações. No entanto, com o tempo, esse coeficiente se tornou obsoleto e precisa ser atualizado para que a empresa possa garantir uma margem mais precisa e eficiente.

Nesse sentido, a empresa está investindo em inteligência artificial para aprimorar seus processos e garantir uma atualização mais precisa do fator de custo. O projeto visa atualizar essa margem de maneira mais assertiva, utilizando ferramentas tecnológicas e dados para criar um modelo mais preciso e eficiente para lidar com as variações de custos e, assim, manter a qualidade e a personalização dos produtos. Com isso, a Tenaris busca se manter na vanguarda do mercado e oferecer a melhor experiência possível aos seus clientes.

BUSINESS UNDERSTANDING

1.3.3 Métrica de sucesso

Com o intuito de garantir o sucesso do projeto, a equipe definiu o desvio médio entre o custo real e o custo estimado pelo fator de custo como a principal métrica a ser considerada. Através dessa medida, será possível avaliar a eficácia do projeto em reduzir as oscilações nos custos dos produtos e minimizar os impactos financeiros para a empresa e seus clientes. Outra métrica a ser utilizada para medir o sucesso pode ser o aumento da eficiência operacional: o uso de um modelo de ML para otimizar o consumo de insumos pode levar a uma maior eficiência operacional, reduzindo o desperdício e os custos associados. Isso pode ser medido pela redução dos custos dos insumos e pela melhoria da produção. A empresa pode utilizar o ROI para estas métricas. Segundo o nosso sponsor, o desvio médio atual gira em torno de 3% a 4%. Sendo assim, o nosso projeto tem como objetivo reduzir ao máximo esse desvio, buscando sempre aprimorar o coeficiente de fator de custo para que ele se torne mais preciso e assertivo.

Além disso, é importante que a equipe trabalhe com o Sponsor e os demais stakeholders do projeto para definir métricas claras de sucesso e monitorar regularmente o progresso. Isso permitirá que a equipe identifique rapidamente quaisquer desvios e ajuste a estratégia conforme necessário para garantir que o projeto atinja seus objetivos.

Com isso, espera-se que o projeto possa trazer benefícios significativos para a empresa, como uma maior eficiência na gestão de custos e uma melhoria na experiência do cliente.

BUSINESS UNDERSTANDING

1.3.4 Problemas da mineração de dados

Um dos principais desafios da mineração de dados é a qualidade dos dados. Os dados podem conter erros, inconsistências e lacunas, o que pode levar a conclusões equivocadas. Além disso, a mineração de dados pode levar a um grande volume de dados irrelevantes ou duplicados, o que pode dificultar a análise e a tomada de decisões.

Outro problema comum é a privacidade e a segurança dos dados. Com a quantidade crescente de dados disponíveis, é cada vez mais importante garantir a privacidade dos dados dos clientes. Além disso, a interpretação dos dados pode ser um desafio, já que os resultados da mineração de dados podem ser complexos e difíceis de entender. Outro problema pode ser na quantidade de dados disponibilizados pela Tenaris, ainda não se sabe ao certo quais dados realmente se terá acesso e se teremos que buscar em outras fontes mais restritivas. Minerar dados requer destreza, tempo, treinamento e visão clara do que se quer alcançar. Neste projeto, o principal facilitador é a disponibilidade e a parceria com o Sponsor, o que vai ajudar na mineração de dados com mais qualidade.

BUSINESS UNDERSTANDING

1.4 PLANO DO PROJETO

1.4.1 Entendimento do negócio

- Identificar os objetivos de negócios do projeto
- Definir as necessidades de análise de dados da empresa
- Identificar as partes interessadas do projeto

1.4.2 Entendimento dos dados

- Identificar as fontes de dados relevantes
- Coletar e analisar os dados
- Identificar problemas de qualidade de dados e tratar as anomalias

1.4.3 Preparação dos dados

- Selecionar os dados relevantes para o projeto
- Transformar, limpar e tratar os dados
- Selecionar os recursos de dados necessários para a modelagem

1.4.4 Modelagem

- Selecionar as técnicas de modelagem mais apropriadas
- Desenvolver modelos de análise de dados
- Avaliar a precisão dos modelos e ajustá-los conforme necessário

1.4.5 Avaliação

- Revisar as descobertas e análises do modelo
- Identificar as ações que precisam ser tomadas para alcançar os objetivos do projeto
- Avaliar o impacto potencial das descobertas do modelo

1.4.6 Implantação

- Desenvolver um plano de ação baseado nas descobertas do modelo
- Implementar as ações recomendadas
- Monitorar continuamente a eficácia das ações tomadas e fazer ajustes conforme necessário

BUSINESS UNDERSTANDING

1.4.7 Tempo

O tempo previsto para a execução de cada etapa é de 15 dias, sendo que algumas atividades serão desenvolvidas em paralelo com outras.

1.4.8 Recursos

Os recursos humanos utilizados, serão os membros da equipe de análise do Grupo de 19 – DS7. Também contaremos com a participação ativa do Sponsor que já se colocou à disposição. Os Recursos materiais e tecnológicos, serão de cada membro da equipe e utilizará softwares gratuitos na sua maioria ou de avaliação trial. Até o momento não há previsão de recursos financeiros. Os riscos associados ao planejamento e execução do projeto, já foram mencionados no item 1.2.3 deste relatório.

As datas que estarão atreladas a este planejamento, serão analisadas e distribuídas em tabela, quando dispomos de todos os dados necessários para início da execução do projeto.