FORMAÇÃO EM DADOS

Modelo de predição de custo de retrabalho

Logotipo

Descrição gerada automaticamente

GRUPO 02

Sumário

[*Modelo de predição de custo de retrabalho* 1](#_Toc183195745)

[1 Business Understanding 3](#_Toc183195746)

[1.1 Objetivo do Negócio 3](#_Toc183195747)

[1.2 Avaliação da Situação 3](#_Toc183195748)

[1.3 Metas de Mineração de Dados 4](#_Toc183195749)

[1.4 Plano do Projeto 4](#_Toc183195750)

[2 Data Understanding 6](#_Toc183195751)

[2.1 Coleta Inicial dos Dados 6](#_Toc183195752)

[2.2 Descrição dos Dados 6](#_Toc183195753)

[2.3 Explorando os Dados 7](#_Toc183195754)

[2.4 Qualidade dos Dados 8](#_Toc183195755)

[3. Data Preparation 9](#_Toc183195756)

### 1 Business Understanding

### **1.1 Objetivo do Negócio**

O objetivo principal deste projeto é identificar maneiras de melhorar a eficiência operacional e a competitividade da Ilumi no setor de materiais elétricos, mitigando problemas como retrabalho e aumentando a qualidade dos produtos para atender às expectativas de consumidores e parceiros de negócios. Além disso, busca-se compreender o impacto das tendências de automação residencial e sustentabilidade na demanda por produtos, oferecendo suporte à tomada de decisão estratégica.

A análise foca em questões-chave, como:

* **Problema:** O retrabalho é uma preocupação significativa, gerando custos elevados e impacto na produtividade.
* **Indicadores e Métricas de Sucesso:** Redução nos custos operacionais associados ao retrabalho, aumento da eficiência produtiva, crescimento na participação de mercado e fortalecimento da imagem da marca no setor.
* **Contexto de Negócio:** O mercado é altamente competitivo, com desafios como escassez de semicondutores e a necessidade constante de inovação. A Ilumi, com sua origem 100% brasileira e portfólio robusto, busca se diferenciar por meio da qualidade, inovação e foco em automação e sustentabilidade.

### **1.2 Avaliação da Situação**

#### Contexto do Setor e da Empresa

O setor de materiais elétricos no Brasil é dinâmico, abrangendo mercados diversificados, como construção civil e automotivo. A Ilumi destaca-se por sua produção de mais de 6 milhões de peças mensais e portfólio com mais de 1.500 itens. No entanto, enfrenta desafios como dependência de insumos importados e presença internacional limitada.

#### Análise SWOT

* **Forças:** Portfólio diversificado, foco em inovação, alta capacidade produtiva e profundo conhecimento do mercado brasileiro.
* **Fraquezas:** Dependência de semicondutores importados e alcance internacional restrito.
* **Oportunidades:** Crescimento da demanda por automação residencial, produtos eco-friendly e expansão para mercados latino-americanos.
* **Ameaças:** Concorrência acirrada, mudanças regulatórias e flutuações econômicas no Brasil.

#### **Observações Importantes**

1. **Defeitos de Componentes:** Estabelecer uma rotina de análise mais rigorosa para componentes adquiridos de terceiros.
2. **Controle de Qualidade:** Iniciar uma pesquisa de mercado para soluções tecnológicas que possibilitem inspeção de 100% das peças.
3. **Monitoramento de Produtos:** Acompanhar o desempenho dos 20 produtos focais e registrar dados para revisões periódicas de orçamento.

#### Cronograma e Riscos

* **Prazo estimado:** De 03/11/2024 a 18/12/2024.
* **Riscos identificados:** Limitações no acesso a dados de alta qualidade, impactos de regulamentações inesperadas e restrições logísticas devido à escassez global de semicondutores.
* **Planos de contingência:** Alternativas para fontes de dados, colaboração com stakeholders e uso de simulações para avaliar cenários.

#### Benefícios Monetários

Reduzir o retrabalho pode trazer economia significativa, aumentar a eficiência produtiva e melhorar a percepção da marca, resultando em maior receita e fidelização dos clientes.

### **1.3 Metas de Mineração de Dados**

A meta técnica do projeto é utilizar análise de dados para identificar padrões e variáveis que contribuem para o retrabalho, além de avaliar tendências de mercado que afetam diretamente a demanda pelos produtos da Ilumi.

#### Critérios de sucesso:

* Desenvolvimento de modelos preditivos para minimizar falhas na produção.
* Identificação de insights para a criação de novos produtos voltados para automação e sustentabilidade.

Justificativa: Essas métricas permitirão alinhar as estratégias de negócios com as demandas do mercado e reduzir custos operacionais.

### **1.4 Plano do Projeto**

#### Tarefas Principais:

* Coleta e exploração de dados internos e externos, incluindo feedback de clientes e desempenho produtivo.
* Análise da qualidade dos dados e tratamento de anomalias.
* Desenvolvimento de modelos analíticos para identificar causas do retrabalho e prever tendências de mercado.
* Apresentação de insights acionáveis para equipes de produção e desenvolvimento de produto.

#### Ferramentas e Tecnologias:

* **Ferramentas de análise:** Python (Pandas, Numpy, Scikit-learn), Excel, Power BI.
* **Técnicas de mineração:** Análise exploratória, modelagem preditiva e visualização de dados.
* **Plataformas:** Material disponibilizado pela empresa e acesso a dados de mercado.

#### Cronograma Resumido:

1. **Fase 1 – Coleta e Compreensão do Problema (03/11 a 10/11):**
   * Coleta de dados internos e externos.
   * Reunião com stakeholders e alinhamento sobre os objetivos do negócio.
2. **Fase 2 – Análise Exploratória e Validação dos Dados (11/11 a 24/11):**
   * Exploração inicial dos dados para identificar padrões e anomalias.
   * Tratamento de dados inconsistentes ou incompletos.
   * Validação da qualidade dos dados para garantir confiabilidade.
3. **Fase 3 – Desenvolvimento de Modelos Analíticos (25/11 a 08/12):**
   * Criação de modelos preditivos para entender os fatores que impactam o retrabalho.
   * Simulações e testes para avaliar possíveis cenários de otimização.
   * Avaliação das tendências de mercado para identificação de novas oportunidades.
4. **Fase 4 – Geração de Relatórios e Recomendações (09/12 a 18/12):**
   * Preparação de um relatório detalhado com insights e recomendações práticas.
   * Criação de visualizações de dados no Power BI para facilitar a comunicação dos resultados.
   * Apresentação final para a banca avaliadora.

### 2 Data Understanding

### **2.1 Coleta Inicial dos Dados**

A coleta inicial de dados foi realizada a partir dos registros de produção, custo, devoluções e defeitos de fabricação das máquinas envolvidas no processo. Os dados foram extraídos de diversas fontes, incluindo planilhas de produção, registros de devoluções e relatórios de qualidade, com informações sobre os tipos de defeitos dos produtos.

#### Fontes de Dados:

* **Planilhas de Produção:** Contêm informações detalhadas sobre a quantidade produzida, custo de produção, tempo de operação e inatividade das máquinas.
* **Registros de Devoluções:** Informações sobre as devoluções dos produtos, incluindo os motivos (defeitos de fabricação, erro de cor, etc.) e a quantidade de unidades devolvidas.
* **Relatórios de Qualidade:** Dados sobre os defeitos encontrados nos produtos, organizados por tipo (defeito de fabricação, defeito de cor, etc.) e correlacionados com a linha de produção e tipo de máquina.

#### Problemas na Coleta de Dados:

* **Dados Incompletos:** Algumas entradas de produção e devoluções não possuem valores registrados para todas as variáveis, como "quantidade total" ou "custo total".
* **Falhas de Registro:** Alguns períodos de inatividade das máquinas não foram registrados corretamente, dificultando a análise de eficiência.

#### Ações para Melhorar a Coleta:

* **Padronização:** Padronizar o processo de coleta de dados para garantir que todas as informações necessárias sejam registradas de forma consistente.
* **Monitoramento em Tempo Real:** Implementar sistemas de monitoramento em tempo real para reduzir erros de registro manual.

### **2.2 Descrição dos Dados**

A descrição dos dados fornece uma visão geral sobre os tipos de informações coletadas e sua organização. Abaixo estão os principais atributos identificados nas fontes de dados.

#### Tabelas e Atributos:

* **Tabela de Produção:**
  + **Máquina:** Identificador da máquina (ex.: Máquina Modelo MT, Máquina Modelo MC).
  + **Data:** Data de registro da produção.
  + **Quantidade Produzida:** Quantidade de unidades produzidas em um determinado período.
  + **Custo de Produção:** Custo total associado à produção, incluindo mão-de-obra e matérias-primas.
  + **Tempo de Inatividade:** Tempo total que a máquina ficou fora de operação, seja por manutenção planejada ou falhas não planejadas.
* **Tabela de Devoluções:**
  + **Produto:** Identificação do produto (ex.: MÓDULO TOMADA 10A PT VIVAZ).
  + **Motivo da Devolução:** Tipo de defeito ou erro (ex.: defeito de fabricação, erro de cor).
  + **Quantidade Devolvida:** Número de unidades devolvidas por tipo de defeito.
  + **Data de Devolução:** Data de registro da devolução.
* **Tabela de Qualidade:**
  + **Produto:** Identificação do produto.
  + **Tipo de Defeito:** Tipo do defeito encontrado (ex.: defeito de cor, defeito mecânico).
  + **Cor do Produto:** Cor do produto no momento da fabricação.
  + **Taxa de Defeito:** Percentual de unidades defeituosas dentro do lote.

#### Estrutura dos Dados:

* **Formato:** As informações são armazenadas principalmente em planilhas Excel, com campos bem definidos, mas com algumas lacunas nos dados.
* **Tipos de Dados:** A maioria dos dados são numéricos (quantidade, custo, tempo de inatividade), enquanto outras variáveis são categóricas (tipo de defeito, cor do produto).

*Possíveis Desafios:*

* **Inconsistência na Coleta de Dados:** Alguns campos estão em formatos diferentes em registros distintos.
* **Falta de Padronização:** A terminologia utilizada para descrever defeitos e cores de produtos não é consistente, o que pode levar a ambiguidades nas análises.

### **2.3 Explorando os Dados**

Na exploração dos dados, o objetivo é identificar padrões, correlações e insights que possam ser úteis para a análise detalhada. Abaixo estão as principais observações encontradas até agora.

*Eficiência de Produção:*

* **Máquinas Modelo MT e MC:** As máquinas possuem um desempenho variável, com a Máquina Modelo MT alcançando uma taxa de produção de 80-82%, enquanto a Máquina Modelo MC está com uma taxa de produção de apenas 50%.
  + **Máquina Modelo MT:** A produção está relativamente estável, com picos de inatividade sendo observados em semanas específicas. Esses picos podem ser explicados por manutenções não programadas ou erros operacionais.
  + **Máquina Modelo MC:** Apresenta uma baixa taxa de produção devido a um número insuficiente de estações de montagem, o que reduz a capacidade de operação.

*Análise de Inatividade:*

A análise das máquinas Modelo MT e MC revela que a inatividade tem um impacto direto na produtividade. Para a Máquina Modelo MT, a inatividade não planejada representa uma redução de até 10% na produção semanal.

*Devoluções de Produtos:*

* **MÓDULO TOMADA 10A PT VIVAZ:** Este produto apresenta uma taxa de devolução elevada, principalmente devido a defeitos de fabricação e erros de cor.

*Correlação entre Defeitos e Cor:*

* Defeitos de Cor: Os defeitos relacionados à cor têm impacto significativo nas taxas de devolução. Este tipo de defeito é mais prevalente em produtos da linha **MÓDULO TOMADA 10A PT VIVAZ**, sugerindo que os processos de pintura ou fabricação podem ser melhorados para evitar esses erros.

*Taxa de Defeito por Produto:*

* **MÓDULO TOMADA 10A SLIM:** Apresenta uma taxa de defeito significativamente menor quando comparado ao **MÓDULO TOMADA 10A PT VIVAZ**, o que pode indicar uma produção mais estável ou controle de qualidade mais eficaz.

*Conclusões da Exploração:*

* **Melhoria da Produção:** A eficiência das máquinas pode ser aumentada com melhor monitoramento da inatividade e implementação de manutenções preventivas.
* **Redução de Devoluções:** Melhorar o controle de qualidade, especialmente em relação aos defeitos de cor, pode reduzir as devoluções e melhorar a satisfação do cliente.

### **2.4 Qualidade dos Dados**

A qualidade dos dados é crucial para garantir a precisão e a confiabilidade das análises. Durante o processo de exploração, foi identificado que os dados apresentam algumas lacunas e inconsistências, que devem ser tratadas para uma análise mais robusta.

*Problemas de Qualidade Identificados:*

* **Valores Faltantes:** Alguns campos importantes, como "custo de produção" e "tempo de inatividade", apresentam valores ausentes, o que pode impactar as análises quantitativas.
* **Inconsistência de Formatação:** Alguns registros de defeitos e cores de produtos estão em formatos diferentes (ex.: cor “branco” e “Branco” sendo tratados como valores diferentes).
* **Erros de Registro:** Falhas de registro, especialmente nos dados de devoluções e na entrada de dados de inatividade, podem afetar a precisão das análises.

*Soluções Propostas:*

* **Tratamento de Dados Faltantes:** Substituir valores ausentes por médias, medianas ou valores estimados, dependendo do tipo de dado.
* **Padronização de Terminologia:** Estabelecer um dicionário de dados para garantir a uniformidade na nomenclatura e formatação dos dados.
* **Verificação de Dados:** Implementar processos de verificação e validação de dados no momento da coleta para reduzir os erros manuais.

### 3. Data Preparation