



MINISTÉRIO DA EDUCAÇÃO  
Universidade Federal de Ouro Preto  
Instituto de Ciências Exatas e Aplicadas  
Especialização em Ciência de Dados



# **O uso da ciência de dados como ferramenta na tomada decisão para análise de custos de celulose na indústria**

**Bruno Oliveira Carvalho**

Belo Oriente, MG  
2024

Bruno Oliveira Carvalho

**O uso da ciência de dados como ferramenta na tomada decisão  
para análise de custos de celulose na indústria**

Trabalho de conclusão de curso apresentado ao curso de Ciência de Dados do Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto, como parte dos requisitos necessários para a obtenção do título de Especialista em Ciência de Dados.

Orientador: Prof. Dr. Thiago Augusto de Oliveira Silva

Belo Oriente, MG

2024

## SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

C331u Carvalho, Bruno Oliveira.

O uso da ciência de dados como ferramenta na tomada decisão para análise de custos de celulose na indústria. [manuscrito] / Bruno Oliveira Carvalho. - 2024.

36 f.: il.: color..

Orientador: Prof. Dr. Thiago Augusto de Oliveira Silva.  
Produção Científica (Especialização). Universidade Federal de Ouro Preto. Departamento de Engenharia de Produção.

1. Celulose - Custo industrial. 2. Celulose - Produção - Estudos de viabilidade. 3. Controle de custo. 4. Indústria de celulose - Previsão - Preços. 5. Processo decisório. I. Silva, Thiago Augusto de Oliveira. II. Universidade Federal de Ouro Preto. III. Título.

CDU 519.2:657.471.1

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431



## FOLHA DE APROVAÇÃO

**Bruno Oliveira Carvalho**

### **O uso da ciência de dados como ferramenta na tomada decisão para análise de custos na indústria de celulose**

Trabalho de conclusão de curso apresentado ao curso de Especialização em Ciência de Dados da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Especialista em Ciência de Dados

Aprovada em 30 de julho de 2024

#### Membros da banca

Dr. Thiago Augusto de Oliveira Silva - Orientador - Universidade Federal de Ouro Preto  
Dr. Alexandre Xavier Martins - Universidade Federal de Ouro Preto  
Me. Ronaldo Neves Ribeiro - Celulose Nipo Brasileira - CENIBRA

Thiago Augusto de Oliveira Silva, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 01/08/2024



Documento assinado eletronicamente por **Thiago Augusto de Oliveira Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 12/08/2024, às 15:48, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0750944** e o código CRC **D55769EF**.

*Este trabalho é dedicado a equipe com a qual trabalho e  
visa aperfeiçoar o desempenho de nossas atividades.*

# Agradecimentos

Agradeço, primeiramente, à Deus!

Agradeço a empresa, que me direcionou a esta linha de aprendizado, tornando possível a abertura deste novo horizonte.

Agradeço à minha família que entendeu os momentos de ausência para os estudos, sem esta compreensão teria sido muito mais difícil.

À UFOP e todos os professores que repassaram seus conhecimentos da melhor forma para esta capacitação.

A todos que de alguma forma contribuíram com este trabalho.

*"Não se pode ensinar nada a um homem.  
Só é possível ajudá-lo a encontrar a coisa dentro de si."  
Galileu Galilei.*

# Resumo

Este trabalho tem o objetivo de investigar e analisar os impactos positivos da aplicação da ciência de dados na qualidade do processo de orçamento, especificamente na avaliação de custos de produção de celulose na indústria. Foi adotada abordagem de pesquisa combinando métodos quantitativos, empíricos e prescritivos, com o propósito de obter uma compreensão profunda da dinâmica dos custos, identificando, classificando e avaliando os gastos envolvidos de forma a compreender os preços dos insumos e suas influências sobre o custo de produção previsto. O processo envolve etapas como pré-processamento de dados, treinamento de modelos (LSTM), avaliação de desempenho e análise financeira. A aplicação dos algoritmos permite previsões precisas de custos e introspecções orientadas por dados. Vale ressaltar a importância dos cenários obtidos, que desempenham um papel fundamental na reflexão crítica e análise da viabilidade para elaboração de futuros orçamentos. Além disso, as informações obtidas podem contribuir de forma significativa para a melhoria contínua do processo de orçamento, auxiliando na tomada de decisões gerenciais com melhores condições de negociação.

**Palavras-chaves:** Orçamento. Custos. Ciência de dados. Informação gerencial.

# Abstract

This work aims to investigate and analyze the positive impacts of the application of data science on the quality of the budgeting process, specifically in the evaluation of pulp production costs in the industry. A research approach combining quantitative, empirical and prescriptive methods was adopted, with the purpose of obtaining a deep understanding of the dynamics of costs, identifying, classifying and evaluating the expenses involved in order to understand the prices of inputs and their influences on the expected cost of production. The process involves steps such as data pre-processing, model training (LSTM), performance evaluation, and financial analysis. The application of algorithms allows for accurate cost forecasts and data-driven insights. It is worth mentioning the importance of the scenarios obtained, which play a fundamental role in critical reflection and feasibility analysis for the preparation of future budgets. In addition, the information obtained can contribute significantly to the continuous improvement of the budget process, helping to make managerial decisions with better negotiation conditions.

**Keywords:** Budget. Costs. Data Science. Management Information.

# Lista de ilustrações

Figura 1 – Função de ativação tanh (Tangente Hiperbólica) . . . . .	7
Figura 2 – Função de ativação Sigmoide. . . . .	7
Figura 3 – Módulo de repetição de uma LSTM. . . . .	8
Figura 4 – Fluxograma . . . . .	9
Figura 5 – Série de dados coletada. . . . .	12
Figura 6 – Melhor modelo . . . . .	15
Figura 7 – Ajustes para o melhor modelo . . . . .	16
Figura 8 – Erro médio por material . . . . .	16
Figura 9 – Material 26 - Menor Erro Médio . . . . .	17
Figura 10 – Material 12 - Menor Desvio Padrão . . . . .	17
Figura 11 – Material 56 - Maior Erro Médio . . . . .	18
Figura 12 – Distribuição . . . . .	18
Figura 13 – Preço futuros . . . . .	19
Figura 14 – Preço futuros (26) . . . . .	19
Figura 15 – Preço futuros (12) . . . . .	19
Figura 16 – Preço futuros (56) . . . . .	20
Figura 17 – Preço futuros (10012) . . . . .	20
Figura 18 – Preço futuros (10223) . . . . .	20
Figura 19 – Preço futuros (10272) . . . . .	21
Figura 20 – Gráfico Violino . . . . .	21

# Lista de abreviaturas e siglas

**ADAM** - Adaptive Moment Estimation

**BW** - Business Warehouse

**Colab** - Google Colaboratory

**Dataframe** - Quadro de dados

**Dropout** - Técnica utilizada para melhorar o desempenho e a generalização dos modelos

**LSTM** - Long Short-Term Memory

**MSE** - Mean Squared Error

**Outliers** - Pontos de dados, fora da curva

**SAP** - Software de gestão

**tanh** - Tangente Hiperbólica

**UFOP** - Universidade Federal de Ouro Preto

**VPL** - Valor Presente Líquido

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
<b>1.1</b>	<b>Objetivo geral</b>	<b>2</b>
1.1.1	Objetivos específicos	2
<b>1.2</b>	<b>Organização do Trabalho</b>	<b>3</b>
<b>2</b>	<b>REVISÃO DA LITERATURA</b>	<b>4</b>
<b>2.1</b>	<b>Tomada de Decisão Gerencial</b>	<b>4</b>
<b>2.2</b>	<b>Produção de Celulose</b>	<b>4</b>
<b>2.3</b>	<b>Complexidade dos Custos na Indústria de Celulose</b>	<b>6</b>
<b>2.4</b>	<b>Rede Neural LSTM</b>	<b>7</b>
<b>2.5</b>	<b>Modelo de simulação</b>	<b>8</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>9</b>
<b>3.1</b>	<b>Instalação de pacotes</b>	<b>10</b>
<b>3.2</b>	<b>Importação das Bibliotecas Necessárias</b>	<b>11</b>
<b>3.3</b>	<b>Acessar o Google Drive</b>	<b>11</b>
<b>3.4</b>	<b>Carregar e Analisar os dados</b>	<b>12</b>
<b>3.5</b>	<b>Definição da Arquitetura da rede LSTM</b>	<b>13</b>
<b>3.6</b>	<b>Treinamento, Avaliação do modelo e Análise de erros</b>	<b>14</b>
<b>3.7</b>	<b>Simulação de cenários</b>	<b>14</b>
<b>4</b>	<b>RESULTADOS</b>	<b>15</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>23</b>
	<b>REFERÊNCIAS</b>	<b>24</b>
	<b>ANEXO A – TABELA DE ERROS</b>	<b>26</b>

# 1 Introdução

A análise de custos é uma parte fundamental da gestão financeira de toda empresa. Ela fornece informações fundamentais para a alocação de recursos, definição de preços e avaliação do desempenho econômico e financeiro, gerando conhecimento quanto aos gastos em bens e serviços em determinado período (SANTOS; SCHIMIDT; PINHEIRO, 2006). Devido esta dinâmica, faz-se necessário a utilização da ciência de dados como ferramenta de suporte para acompanhamento e obtenção da lucratividade esperada pelos acionistas, potencialmente proporcionando uma nova abordagem para o desenvolvimento sustentável da economia e do meio ambiente (HEINRICHS, 2013). O problema levantado reside no fato de que a análise de custos tradicional muitas vezes depende de métodos estáticos e históricos que não conseguem capturar a complexidade dos dados financeiros, a exemplo a metodologia descrita segundo (MATZ; CURRY; FRANK, 1978). Os custos podem variar significativamente devido a fatores sazonais, flutuações no mercado, mudanças nas operações da empresa e uma série de outros aspectos, além de ser utilizado em todas as etapas da contabilidade de custos: no planejamento, na execução e na interpretação dos gastos.

A celulose é um dos principais produtos exportados no Brasil. A área estimada de florestas plantadas totalizou, em 2022, 7,3 milhões de hectares, dos quais 70,1% estão concentrados nas regiões Sul e Sudeste. As áreas com cobertura de eucalipto e pinus corresponderam a 96% das florestas plantadas para fins comerciais no país (IBGE, 2022). O Brasil é referência mundial em produtividade de plantios florestais devido às condições de clima e solo, além de décadas de investimentos em pesquisa, desenvolvimento e inovação; técnicas de manejo florestal; melhoramento genético e práticas sustentáveis. O eucalipto e o pinus plantados hoje evoluíram em produtividade 40% a mais que as espécies e variedades trazidas ao Brasil, na sua origem. (EMBRAPA, 2021). Diante destes dados pode ser visto em conjunto o consumo específico de madeira utilizado nas fábricas pode variar muito em função de vários fatores, sendo o principal deles a eficiência da fábrica, o nível tecnológico e o preço dos insumos que impactam de acordo com o tipo/qualidade da madeira exaurida, que pode gerar necessidade de revisão da quantidade de consumo específico de outros insumos para obtenção da qualidade desejada, impactando diretamente nos custos de produção. (QUEIROZ, 2004)

A Ciência de Dados oferece a capacidade de lidar com a complexidade ao empregar algoritmos avançados que identificam padrões complexos e tendências sutis nos dados, impedindo a alocação excessiva de recursos e evitando danos à saúde financeira da organização. (TURBAN *et al.*, 2011) A integração da Ciência de Dados na análise de custos empresariais não é apenas relevante, mas também imperativa para as organizações que buscam se manter competitivas e bem-sucedidas em um cenário empresarial em constante evolução.

Nesse contexto, este trabalho busca responder como e quais os modelos de aprendizado de máquina podem ser treinados para fazer previsões de tendências de custos futuros, permitindo que as empresas se preparem adequadamente e tomem decisões informadas. (RAUTENBERG SANDRO; CARMO, 2019) Isso não apenas melhora a rentabilidade, mas também ajuda a empresa a permanecer competitiva em seu mercado, além de contribuir para a detecção de erros financeiros.

## 1.1 Objetivo geral

O objetivo geral deste trabalho é analisar os custos de produção de celulose na indústria com o uso da ciência de dados, visando melhorar a qualidade do processo orçamentário com previsão de preço de insumos para facilitar a tomada de decisão tornando-a baseada em dados.

### 1.1.1 Objetivos específicos

Para cumprimento do objetivo geral é necessário atender aos seguintes objetivos específicos:

- Utilizar a arquitetura de rede neural LSTM para gerar simulações de cenários;
- Analisar a distribuição de erros de previsão dos modelos;
- Identificar variáveis que influenciam os custos;
- Calcular a previsão de preços futuros para os custos envolvidos no processo de fabricação da celulose;
- Identificar através dos resultados sugestões de melhoria.

## 1.2 Organização do Trabalho

Este trabalho está organizado em cinco capítulos. No Capítulo 1 os assuntos de análise de custos, produção de celulose e ciência de dados são introduzidos. A Revisão Bibliográfica, dividida em quatro seções no Capítulo 2. A primeira seção apresenta uma breve definição de tomada de decisão gerencial, a segunda e terceira descrevem o processo de produção de celulose, incluindo a complexidade dos custos de insumos básicos para a elaboração de cenários e as tecnologias utilizadas. Na Seção 2.4, o modelo de rede LSTM é fundamentado, e em seguida, abordado o modelo de simulação responsável por gerar os novos preços, tornando possível a análise financeira dos Valores Presentes Líquidos. No Capítulo 3 apresenta a metodologia aplicada. Os resultados obtidos a partir da aplicação da simulação da rede neural LSTM em um conjunto de dados dos últimos 5 anos no ambiente *Python* são apresentados no Capítulo 4. O Capítulo 5 apresenta as considerações finais do trabalho, limitações e possíveis trabalhos futuros.

## 2 Revisão da Literatura

### 2.1 Tomada de Decisão Gerencial

No contexto da tomada de decisão gerencial, a informação se configura como um processo crucial para a navegação em um mar de incertezas e a conquista de resultados expressivos. Guiar equipes, alocar recursos e direcionar estratégias exigem dos gestores um domínio ímpar da arte de decidir. E nesse cenário, (BAZERMAN M. H.; MOORE, 2017) exploram como os gestores tomam decisões e os desafios que enfrentam devido a vieses cognitivos. Discutindo também estratégias para melhorar a tomada de decisão, como a promoção da consciência sobre esses vieses e a implementação de práticas baseadas em evidências.

A ciência de dados, em conjunto com a abordagem baseada em simulações, representa uma transformação poderosa na forma como as organizações tomam decisões. Ao explorar o potencial dessa ferramenta, as empresas podem navegar com mais segurança em um mundo de incertezas, aprimorar seu desempenho, reduzir riscos, fomentar a inovação e alcançar um sucesso sustentável no longo prazo. A conclusão a que podemos chegar é a frase de Peter Drucker, pai da administração moderna: “O planejamento não diz respeito a decisões futuras, mas às implicações futuras de decisões presentes”.

### 2.2 Produção de Celulose

A produção de celulose é um processo que envolve diversas etapas, desde a preparação das mudas, a colheita da matéria-prima até a obtenção da produção da celulose. A produção de celulose, desde a floresta até o produto final, é um processo complexo e fascinante que envolve diversas etapas interligadas onde cada etapa contribui para a transformação da madeira em celulose, um material essencial para diversas utilidades na sociedade. Abaixo, são descritas as principais etapas desse processo (CAMPOS, 2008):

A primeira etapa do processo inicia-se com o preparo das mudas, que é o ponto de partida para o crescimento das árvores, estas mudas vêm de uma seleção rigorosa de sementes/clones de eucalipto ou pinus para garantir o desenvolvimento saudável das árvores ao longo do ciclo produtivo.

No plantio e manejo florestal, o uso de adubos e fertilizantes ricos em nutrientes são essenciais para nutrir as árvores e promover seu crescimento. Ademais, a aplicação criteriosa de agrotóxicos e inseticidas seletivos é fundamental para o controle de pragas, doenças e ervas daninhas que ameaçam as florestas. A gestão eficiente da água, por sua vez, é vital para o desenvolvimento das árvores, desde a irrigação das mudas até o manejo das florestas plantadas, garantindo a otimização desse recurso precioso e a preservação dos recursos hídricos.

Na etapa de colheita e transporte, o uso de combustíveis renováveis ou de alta eficiência energética contribui para a redução das emissões de gases de efeito estufa e para a minimização do impacto ambiental. Além disso, a realização de manutenções preventivas regulares são fundamentais para garantir a eficiência e a segurança das operações durante a colheita e o transporte da madeira.

Por fim, no processo industrial de transformação da madeira em celulose, a qualidade da madeira proveniente de florestas manejadas de forma sustentável é crucial para a eficiência do processo e a qualidade da celulose produzida. A escolha adequada do licor químico para o cozimento da madeira e dos produtos químicos para o branqueamento da polpa de celulose é essencial para garantir a qualidade do produto final e minimizar o impacto ambiental. Além disso, o uso de fontes de energia renováveis ou de alta eficiência energética contribui para a redução das emissões de gases de efeito estufa e para a minimização do impacto ambiental durante todo o processo produtivo.

Ao longo do processo de produção de celulose, a utilização criteriosa de insumos específicos em cada etapa é fundamental para garantir a qualidade, a eficiência e a sustentabilidade da produção. A seleção cuidadosa desses insumos e sua utilização eficiente são fundamentais para o desenvolvimento de um material essencial para diversos setores, demonstrando o compromisso da indústria de celulose com a inovação, a qualidade e a responsabilidade ambiental.

A simulação de preços da madeira, produtos químicos, energia elétrica, gás natural e outros combustíveis, representam a composição dos custos necessários a fabricação da celulose, esta simulação visa evidenciar o ganho na alocação eficiente de recursos otimizando o desempenho. (CARVALHO, 2006) Explorar avanços tecnológicos, quanto a inovações na produção de celulose, o uso de inteligência artificial e tecnologias sustentáveis serão detectados na eficiência operacional e na otimização de processos industriais. Algoritmos avançados podem analisar grandes volumes de dados em tempo real, identificando padrões e fornecendo *insights* para melhorar a eficiência operacional.

Na produção de celulose, a inteligência artificial pode ser empregada na previsão de demanda, controle de qualidade e manutenção preditiva de equipamentos, tecnologias sustentáveis na gestão de resíduos têm sido incorporadas para gerenciar resíduos gerados no processo de produção e sistemas de reciclagem de água, por exemplo, reduzem o consumo de recursos hídricos. Além disso, tecnologias de recuperação de produtos químicos contribuem para a minimização do impacto ambiental, promovendo a sustentabilidade.

## 2.3 Complexidade dos Custos na Indústria de Celulose

A indústria de celulose, por sua natureza complexa e dependente de diversos fatores externos, apresenta uma estrutura de custos igualmente complexa. Diversos elementos interagem e influenciam os custos de produção, tornando a gestão financeira desse setor um desafio constante. Segundo (CAMPOS, 2008) os principais fatores que contribuem para a complexidade dos custos são:

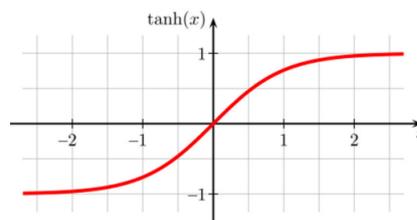
**Matéria-prima** a madeira, principal insumo da indústria, apresenta variações significativas de preço em função de fatores como oferta e demanda, condições climáticas, e políticas florestais. **Energia** a produção de celulose é um processo altamente energético, e as variações nos preços da energia, principalmente da eletricidade e dos combustíveis fósseis, impactam diretamente nos custos de produção. **Produtos químicos** a utilização de produtos químicos no processo de produção da celulose representa uma parcela significativa dos custos, e a volatilidade dos preços desses insumos também influencia os custos finais. **Mão de obra** os custos com mão de obra representam uma parcela importante dos custos totais, e são influenciados por fatores como negociações coletivas, legislação trabalhista e qualificação da mão de obra. **Logística** a logística de transporte da madeira e do produto final representa um custo significativo, especialmente em países com grandes extensões territoriais como o Brasil. **Taxas e impostos** a carga tributária sobre a indústria de celulose é significativa e varia de acordo com a legislação de cada país. **Investimento em tecnologia** A indústria de celulose é caracterizada por um alto investimento em tecnologia, visando aumentar a eficiência e reduzir custos. No entanto, esses investimentos geram custos fixos elevados. **Sustentabilidade** as empresas do setor estão cada vez mais pressionadas a adotar práticas sustentáveis, o que pode gerar custos adicionais relacionados à gestão ambiental e social.

Segundo estudo (SANTOS, 2018), a eficiência florestal no Brasil se destaca globalmente devido ao cultivo eficiente de eucalipto, matéria-prima principal da celulose. Com ciclos de colheita mais curtos e tecnologias avançadas, as empresas reduzem custos e melhoram a competitividade. A integração vertical, reduz custos eliminando etapas intermediárias e maximizando eficiência energética. Destaca-se o consumo de combustíveis, a soda, o ácido sulfúrico e gás natural que são muito voláteis em relação a mudança de preço, podendo gerar inflação de demanda por expectativa para orçamento.

## 2.4 Rede Neural LSTM

Como os seres humanos, as redes neurais recorrentes de Longo Prazo (LSTM) não iniciam seus pensamentos ou processamento do zero. A cada palavra ou dado novo ela gera compreensão em relação à anterior aprendida, também conhecida como *Long Short-Term Memory Networks*, é um tipo especial de rede neural artificial que se destaca por sua capacidade de aprender e processar sequências de dados de forma eficiente, superando as limitações de modelos tradicionais. (LIU; CHAO; LIN Y.-C.AND LIN, 2019; MEHTAB; SEN; DUTTA, 2020) afirmam que essa técnica tem se mostrado eficiente por ser capaz de analisar grandes quantidades de dados históricos, identificar padrões e tendências, onde, as séries temporais apresentam quatro principais componentes, que são capturados ao longo do processamento do modelo sendo: tendência, ciclo, sazonalidade, irregularidade ou fatores aleatórios. Detalhando o módulo LSTM, temos a **FIGURA 1** representando a função de ativação tangente hiperbólica ( $\tanh$ ) que é utilizada para introduzir não-linearidade no modelo. Essa não-linearidade é crucial para que a rede seja capaz de aprender representações complexas dos dados e realizar tarefas como classificação e regressão.

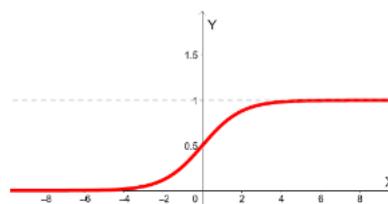
Figura 1 – Função de ativação  $\tanh$  (Tangente Hiperbólica).



Fonte: o Autor.

Na **FIGURA 2** representando a limitação da função sigmoide em uma LSTM mapeando qualquer valor real para um número entre 0 e 1, é como um portão que controla o fluxo de informações dentro da rede. Ela permite que a LSTM decida quais informações são relevantes para a tarefa em questão e quais podem ser descartadas.

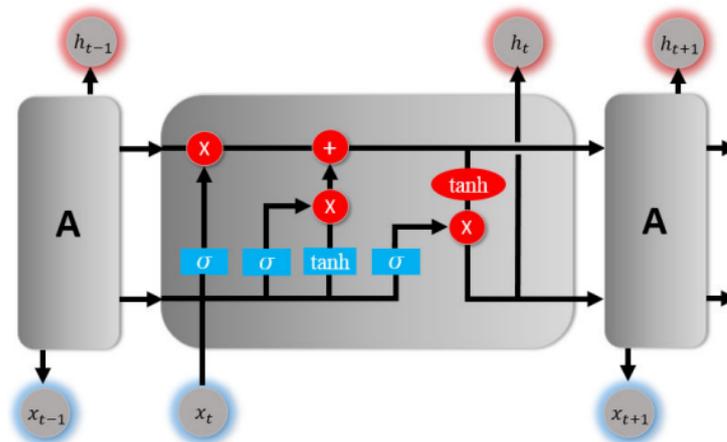
Figura 2 – Função de ativação Sigmoide.



Fonte: o Autor.

Na **FIGURA 3** do módulo LSTM são visualizadas as células de memória e as três portas do módulo, sendo a primeira a porta de esquecimento, onde decide quais informações devem ser descartadas, a segunda porta de entrada que determina quais novas informações devem ser armazenadas na memória e a porta de saída que decide quais informações da célula de memória serão utilizadas para calcular a saída da LSTM.

Figura 3 – Módulo de repetição de uma LSTM.



Fonte: (OLAH, 2015)

## 2.5 Modelo de simulação

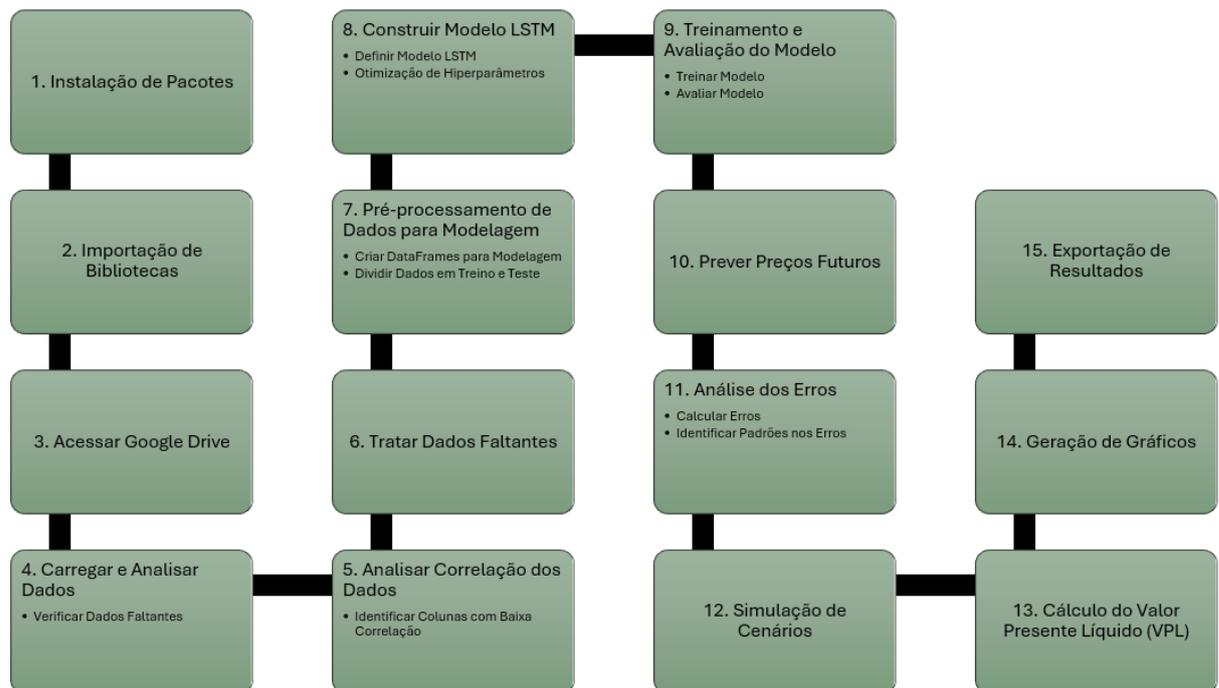
O modelo de simulação desenvolvido tem o objetivo principal de realizar a análise de viabilidade econômica, utilizando previsões de preços de materiais, considerando as quantidades de produção e custo total para o cálculo do valor presente líquido (VPL). O modelo simula diferentes cenários de preços futuros e avalia o impacto desses cenários no resultado financeiro utilizando a análise de viabilidade econômica conforme utilizado por (REZENDE J. L. P.; OLIVEIRA, 2008), como a métrica para análise de projetos florestais.

## 3 Metodologia

Este trabalho foi baseado na utilização do método de pesquisa quantitativa concentrando em dados numéricos e análise estatística para entender e explicar fenômenos gerados no orçamento pelo preço estimado de alguns insumos, onde, o método preocupa-se principalmente em quantificar variáveis, fornecendo informações numéricas que podem ser analisadas objetivamente, a partir da projeção do sistema real em linguagem matemática e computacional (CAUCHICK-MIGUEL *et al.*, 2018), tendo a ideia de que o modelo proposto neste trabalho seja capaz de representar um conjunto de políticas, regras e procedimentos que juntos simulem o ambiente real do estudo.

A **FIGURA 4**, são sintetizados os passos metodológicos necessários para a realização do objetivo do trabalho.

Figura 4 – Fluxograma



Fonte: o Autor.

Neste contexto a pesquisa permite que identifique padrões e relações com base nos dados históricos dos últimos 5 anos. Os dados são coletados e organizados em uma planilha, onde cada coluna representa um material diferente e cada linha representa uma data. Em seguida, os dados são transformados em uma lista de estrutura de dados, onde cada *dataframe* contém duas colunas: "Data" e "Preço". Os insumos necessários para o processamento e obtenção da celulose avaliados neste trabalho estão divididos em 60 produtos diferentes conforme as quantidades necessárias para atendimento do balanceamento projetado da fábrica. Diante das informações históricas torna possível o início da criação de um modelo digital que imita processos do mundo real e permite o teste de diferentes cenários. As simulações no contexto da análise de custos da indústria de celulose fornecem um ambiente controlado para examinar os diferentes cenários de custo e seus efeitos sobre o resultado geral no período determinado.

É oportuno destacar a afirmação de (PADOVEZE, 2012) quanto a representação dos métodos de custeio partindo do total de custos indiretos e diretos associados às simulações como um caminho expressivo para compreender o fenômeno de múltiplos cenários aleatórios para modelar sistemas complexos e fornecer resultados probabilísticos. Conforme comentado por (GENTLE, 2003) ao simular os custos potenciais, estes podem determinar o curso de ação mais eficiente e decisões informadas em relação à gestão de custos, demonstrando a confiabilidade e validade das principais razões para a escolha de um método de simulação.

### 3.1 Instalação de pacotes

A instalação dos pacotes *keras-tuner* e *distfit* no GOOGLE COLAB foram necessárias devido às funcionalidades específicas durante o desenvolvimento do algoritmo sendo, a busca de hiper parâmetros para controle da melhoria de performance do processo de treinamento do modelo e a realização do ajuste de distribuições estatística que melhor ajusta aos dados respectivamente.

## 3.2 Importação das Bibliotecas Necessárias

A importação das bibliotecas é uma etapa fundamental na construção do projeto. Elas fornecem as funcionalidades pré-construídas auxiliando no desenvolvimento dos modelos. Para este trabalho foram necessárias as seguintes bibliotecas: **NumPy** - usada principalmente para manipulação e operações aritméticas eficientes em *arrays* de dados numéricos, **Pandas** - Usada para carregar e preparar os dados, permitindo fácil acesso e manipulação das colunas e linhas dos *datasets*, **Matplotlib** - Usada para criar gráficos e visualizações dos dados, facilitando a interpretação dos resultados, juntamente com a **Seaborn** - Oferecendo uma interface mais alta e intuitiva para criar gráficos estatísticos atraentes e informativos, **TensorFlow/Keras** - Utilizadas para definir e treinar redes neurais, especificamente redes LSTM (*Long Short-Term Memory*), que são adequadas para modelagem de séries temporais, **Keras Tuner** - Usada para realizar a busca de hiper parâmetros, encontrando a melhor combinação de parâmetros para o modelo LSTM, **Scikit-learn** - Usado para normalizar os dados, escalando os valores para um intervalo entre 0 e 1 e a **Distfit** - Usada para ajustar a distribuição dos erros das previsões do modelo, permitindo uma análise detalhada dos desvios e a geração de cenários futuros baseados nesses erros ajustados.

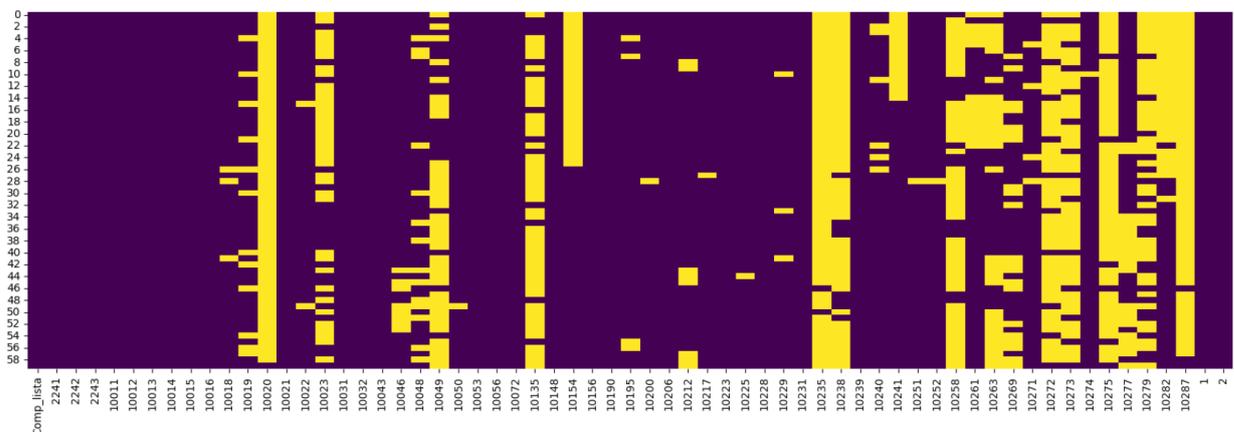
## 3.3 Acessar o Google Drive

Para acesso aos dados eles foram disponibilizados via Google Drive onde para serem lidos necessitam da disponibilização no ambiente de desenvolvimento permitindo que acesse os arquivos e diretórios diretamente no notebook do Colab. Este acesso é realizado através do código: `'drive.mount('/content/drive')`.

### 3.4 Carregar e Analisar os dados

Para o trabalho proposto foram extraídas informações dos preços praticados de 60 materiais diferentes registrados no sistema (SAP) no período dos últimos 5 anos retroativos a 2024, através da ferramenta *Business Warehouse (BW)*, sendo estes dados reais retirados após o processamento das rotinas de custos mensais contendo a identificação dos respectivos materiais e salvos em arquivo em excel. Na **FIGURA 5**, são demonstrados determinados períodos de tempo que não ocorre a necessidade de alguns materiais, isso ocorre conforme o balanço de consumo da fábrica, porém para o uso da simulação em um ambiente controlado para garantir que os fatores mais relevantes sejam considerados e contabilizados torna-se necessário ajustar a base de dados de forma que não possuam dados faltantes, assim, foram utilizados os métodos ( *Forward Fill – ffill* ) da biblioteca Pandas para tratar os valores ausentes completando as linhas anteriores em branco e o método ( *Backward Fill – bfill* ) para preencher as células em branco com o próximo valor válido, observando na mesma coluna após a célula com o valor ausente. Este método foi utilizado visando a possibilidade de generalização com outras bases de dados.

Figura 5 – Série de dados coletada.



Fonte: O Autor.

### 3.5 Definição da Arquitetura da rede LSTM

Nesta etapa, os dados foram normalizados para evitar redundância de informações e garantir que todos os atributos estejam no mesmo intervalo de valores preservando a integridade dos dados, logo, estes são particionados em grupos de treinamento e teste, sendo 80/20 respectivamente para testar e avaliar a performance do modelo de maneira objetiva.

A arquitetura da rede construída foi elaborada criando uma lista de dataframes contendo os dados de cada material específico com Data e preço, logo, ele executa os passos: de Normalização utilizando o **RobustScaler** que calcula a mediana e os quartis dos dados para remover a influência de *outliers*. Em seguida, os dados são escalados de acordo com esses valores; Separação de Conjuntos de treino e teste, gerando previsões para cada dataframe dentro da coleção; Criação de sequências onde os dados são transformados em sequências de tamanho **k** (15 utilizado no modelo) e cada sequência representa os k passos anteriores da série temporal. Por exemplo, se  $k=3$ , a primeira sequência será  $[t-3, t-2, t-1]$ , e a segunda será  $[t-2, t-1, t]$ . Essas sequências são utilizadas na etapa de treinamento como entradas para a LSTM, que aprenderá a mapear as sequências passadas para prever o valor futuro repetindo esse processo para gerar previsões para os próximos 12 meses.

A definição da arquitetura foi proposta utilizando números inteiros entre 32 e 128 com incremento de 16 unidades nas camadas ocultas e após cada camada utiliza o "dropout rate" para valores entre 0 e 0,3 para evitar sobre ajustes. Após duas camadas "LSTM" foi executada uma camada densa considerando além dos parâmetros mencionados a função de ativação "RELU" a "TANH - Tangente hiperbólica" e por fim, uma camada de saída densa com a função de ativação sigmoide. Logo, realiza o treinamento com a função "model.compile" utilizando como otimizador o "ADAM - Adaptive Moment Estimation" porque ele se adapta a taxa de aprendizado para cada parâmetro individualmente, considerando gradientes passados para acelerar a convergência e a função de perda o Erro médio quadrático "MSE - Mean Squared Error" assim, o otimizador Adam ajusta os parâmetros internos do modelo e o "MSE" mede o quão bem as previsões do modelo correspondem aos dados reais.

Um ponto importante que foi considerando é a possibilidade de otimização dos hiper parâmetros, utilizando o "Keras Tunner" foi criada uma instância do *Hyperband tuner* e definida a função objetivo como o melhor resultado do menor "MSE - Erro quadrático médio" na validação modelo passando por no máximo 10 épocas de treinamento em cada modelo e definindo um diretório e nome do projeto a ser registrado, após a busca, recupera o melhor modelo encontrado.

## 3.6 Treinamento, Avaliação do modelo e Análise de erros

Diante do melhor modelo encontrado pela otimização de hiper parâmetros é realizada a busca dos dados originais de treinamento e teste e realiza novas previsões no conjunto de dados para calcular inicialmente o erro e depois calcula a diferença entre os valores reais e os erros encontrados dos novos valores previstos de preço. Porém estes *arrays* encontrados estão na estrutura de 12 x 33 sendo necessário a remodelagem do *array* de erros para ter uma forma específica (396 elementos) para ajuste da densidade de probabilidade de distribuições durante a execução do **distfit** que será citado adiante.

A lista de erros obtida é usada para ajustar distribuições de probabilidade aos dados utilizando o **distfit** que busca a melhor distribuição normal para representar cada conjunto de erros armazenando os parâmetros de média e desvio padrão da distribuição ajustada em uma lista estes dados podem ser observados no **ANEXO A** para avaliação dos erros obtidos durante a geração do modelo, nesta lista apresenta os erros classificados até o quito pior erro além da média e o desvio padrão para todos os materiais.

## 3.7 Simulação de cenários

A simulação foi gerada considerando a incerteza nos preços de diversos ativos e utilizando modelos de previsão para gerar preços futuros e, em seguida, calcular o Valor Presente Líquido (VPL) para diferentes cenários. Tornando possível simular fluxos de caixa futuros e avaliar a incerteza associada a um projeto de investimento ou um orçamento por completo.

A função utilizada percorre cada ativo da lista gerando a previsão de preço que é adicionado um ruído aleatório, seguindo uma distribuição normal média dos erros calculados anteriormente pelo **distfit**. Isso foi feito para adicionar a simulação a incerteza intrínseca do mercado financeiro.

Após o cálculo da lista de preços de insumos, estes foram multiplicados pelos respectivos volumes de consumo específico necessário para atendimento a fábrica, obtendo assim os valores de custo variável dos próximos meses. Este custo variável adicionado o custo fixo mensal trazido diretamente dos coletores de custo preestabelecidos formam a função de "**calc custo**", também foi trazido diretamente os valores de receita dos períodos para complementação do cálculo do fluxo de caixa e o VPL.

Por fim, a função de simulação é chamada para gerar os múltiplos cenários econômicos, calculando seus respectivos VPL's, adicionando durante a execução o parâmetro de horizonte de tempo previsto de "**12 meses**", taxa de desconto de **10%** diante do cenário atual da Selic 10,5% e realizando este procedimento para **100** iterações.

## 4 Resultados

O algoritmo desenvolvido utilizou redes neurais **LSTM** para prever os preços futuros de diferentes materiais com base em dados históricos e a arquitetura utilizada foi ajustada com a ajuda do **Keras Tuner** para otimizar os hiper parâmetros, como o número de neurônios, a função de ativação e a taxa de *dropout*. A base de dados utilizada foi ajustada para preencher valores faltantes e normalizados utilizando o **RobustScaler** logo, o conjunto de dados foi dividido em treino e teste para permitir a avaliação do modelo utilizando métricas como *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE). Estas métricas forneceram informações sobre o desempenho dos modelos em prever os preços futuros dos materiais.

O melhor modelo obtido a partir do processo de *tuning* é recuperado usando "best model = tuner.get\_best\_models()[0]". Na **FIGURA 6** é visualizado a representação da configuração que obteve o menor erro quadrático médio de validação durante a pesquisa, exibindo o modelo descrito como "sequencial" e inclui várias camadas, como camadas LSTM, *dropout* e densas. Cada camada tem sua forma de saída e número de parâmetros listados. Por exemplo, a primeira camada LSTM tem uma forma de saída de (Nenhum, 128) com 65.536 parâmetros. O total de parâmetros é indicado como 2.162.624, que requer aproximadamente 844,55 KB de memória. já a **FIGURA 7** é mostrado como o modelo realiza a busca por melhores parâmetros ajustando a função de ativação e a quantidade de épocas exemplificados nas quatro saídas coletadas durante o processamento.

Figura 6 – Melhor modelo

```
o melhor modelo é
Model: "sequential"
```

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 15, 128)	66560
dropout (Dropout)	(None, 15, 128)	0
lstm_1 (LSTM)	(None, 128)	131584
dropout_1 (Dropout)	(None, 128)	0
dense (Dense)	(None, 128)	16512
dropout_2 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 12)	1548

```

Total params: 216204 (844.55 KB)
Trainable params: 216204 (844.55 KB)
Non-trainable params: 0 (0.00 Byte)

```

Fonte: o Autor.

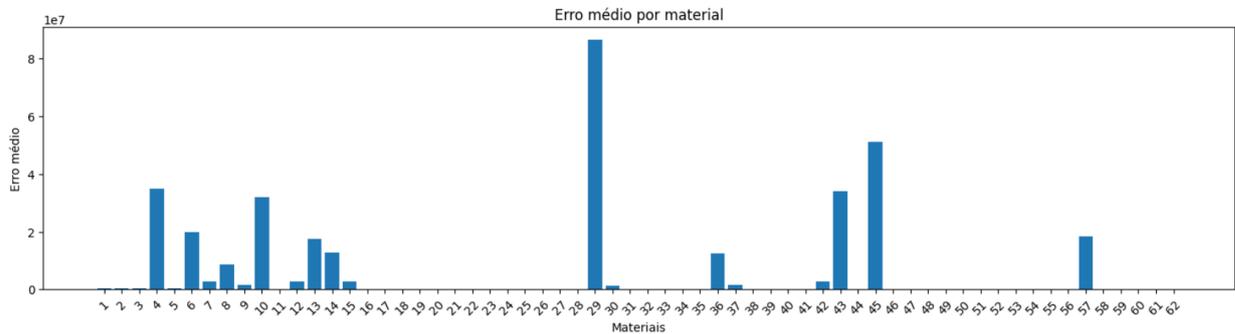
Figura 7 – Ajustes para o melhor modelo

Trial	Trial	Trial	Trial	Trial
summary	28	summary	29	summary
Hyperparameters:		Hyperparameters:	13	Hyperparameters:
units:	112	units:	80	units:
activation:	relu	activation:	112	activation:
dropout:	0.2646074705862067	dropout:	0.1692019497013076	dropout:
tuner/epochs:	10	tuner/epochs:	10	tuner/epochs:
tuner/initial_epoch:	0	tuner/initial_epoch:	0	tuner/initial_epoch:
tuner/bracket:	0	tuner/bracket:	0	tuner/bracket:
tuner/round:	0	tuner/round:	0	tuner/round:
Score:	0.23569142818450928	Score:	0.2547234892845154	Score:
			0.3172871470451355	Score:
			0.3252926766872406	

Fonte: o Autor.

O modelo LSTM com os dados reais, através da coleção de estatísticas de erro médio gerada para cada material apresentado na **FIGURA 8**, é visualizado o comportamento de erro gerado para alguns materiais, isso ocorre em função da oscilação na necessidade destes materiais, que não possuem informações suficientes para geração de bons cenários. No eixo X estão os materiais e no eixo Y estão os valores de erro médio.

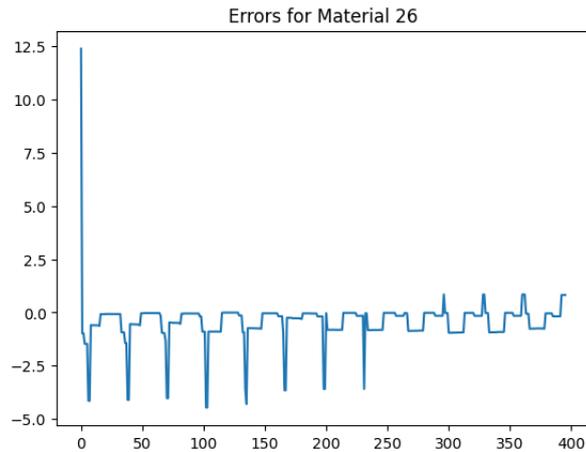
Figura 8 – Erro médio por material



Fonte: o Autor.

O modelo está utilizando uma base de 60 materiais diferentes e torna visível que o comportamento para alguns materiais não atende a expectativa. O processamento da distribuição dos erros de forma normal apresenta os extremos da análise de erros gerados onde, o **Material 26** possui o menor erro médio sugerido e o modelo se comporta bem em média para prever o preço deste material, pois os erros estão mais próximos de zero, conforme **FIGURA 9**. No eixo X estão as elementos e no eixo Y estão os valores de erro médio.

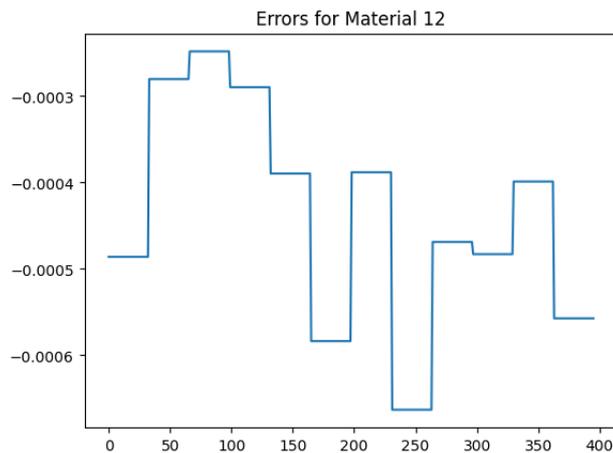
Figura 9 – Material 26 - Menor Erro Médio



Fonte:O Autor.

Na **FIGURA 10**, o **Material 12** é apresentado o menor desvio padrão dos erros. Isso implica que as previsões do modelo para este material são as mais consistentes, com os erros agrupados em torno do valor médio. Há menos variação na precisão das previsões para este material onde, estão menos propensas a se desviar significativamente dos preços reais. No eixo X estão os elementos e no eixo Y estão os valores de erro médio variando entre -0,0006 e -0,0003. O **Material 56** como pode ser visualizado na **FIGURA 11**, indicando que o modelo

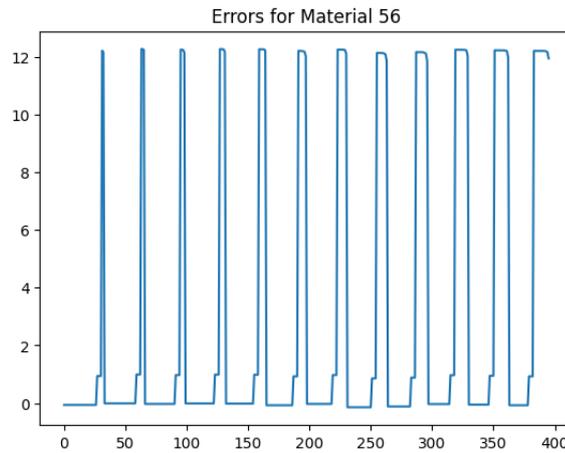
Figura 10 – Material 12 - Menor Desvio Padrão



Fonte:o Autor.

consistentemente subestima ou superestima o preço para este material, e as previsões para este material também são as mais dispersas, significando que possuem a menor precisão. No eixo X estão os elementos e no eixo Y estão os valores de erro médio que chega a atingir o índice 12.

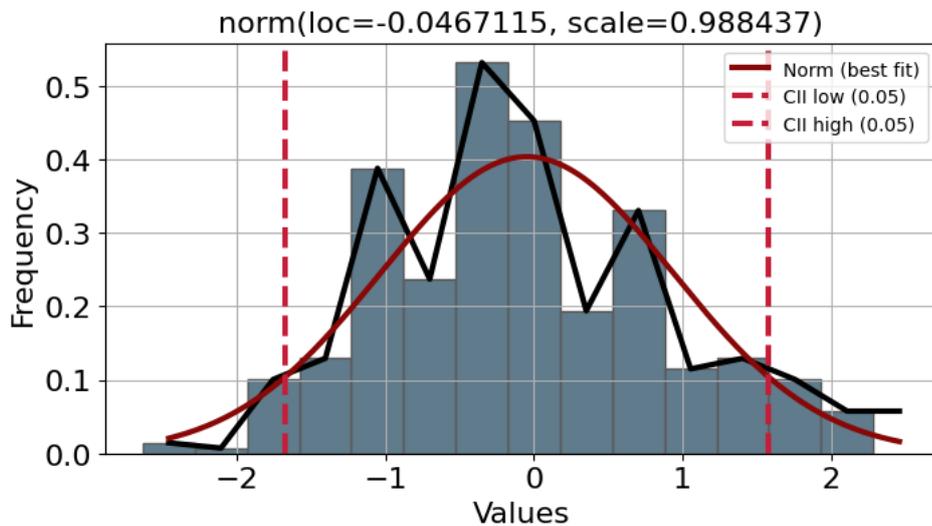
Figura 11 – Material 56 - Maior Erro Médio



Fonte: o Autor.

A FIGURA 12 é exibido um histograma sobreposto com uma curva de função de densidade de probabilidade dos erros na cor vermelha suave, rotulada como “Norm (melhor ajuste)”, ela se ajusta bem ao topo das barras do histograma. Isso sugere que ela representa a distribuição normal que melhor se ajusta aos dados observados e as linhas de intervalo de confiança com nível de significância de 5% representado por linhas verticais tracejadas em vermelho. As barras azuis representam a distribuição de frequência dos pontos dos dados em diferentes faixas de valores no eixo x, que está rotulado como “Valores”. Cada barra indica quantas vezes um valor específico aparece nos dados. O eixo y é rotulado como “Frequência” e varia de 0 a 0,55. Isso significa que a frequência máxima observada nos dados é 0,53. A informação "loc - média" e "scale - desvio padrão".

Figura 12 – Distribuição



Fonte: o Autor.

Após esta etapa, reverte a normalização realizada anteriormente nos dados, convertendo as previsões para a escala original dos preços futuros gerados retornando a identificação dos códigos de cada material, conforme **FIGURA 13**.

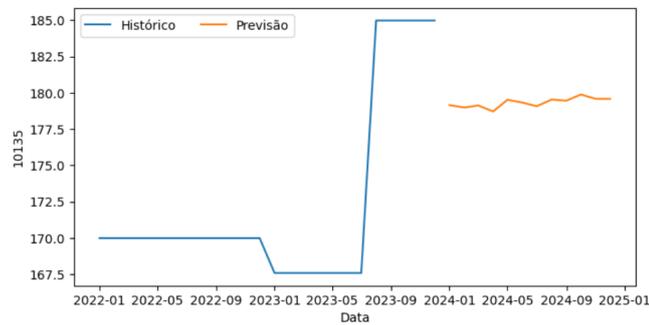
Figura 13 – Preço futuros

Data	2241	2242	2243	10011	10012	10013	10014	10015	10016	10018	...	10272	10273
01/01/2024	548.077393	548.142822	511.198151	5969.718750	555.036011	4512.481934	1637.265869	2997.442139	1314.435913	5719.802246	...	30.931469	41.527569
01/02/2024	544.420166	544.848755	506.181061	5911.975586	549.773743	4478.340332	1620.983032	2971.879150	1296.272095	5642.973633	...	30.700962	41.332581
01/03/2024	547.323120	547.439697	510.138123	5957.919922	553.942383	4504.638672	1632.885742	2991.636475	1310.860474	5704.949707	...	30.871838	41.483322
01/04/2024	527.223145	529.432861	482.674469	5632.964844	524.888428	4324.702148	1546.799927	2852.592285	1209.624756	5265.357422	...	29.685215	40.418846
01/05/2024	549.515259	549.644287	512.545288	5987.028809	557.864441	4519.966309	1640.021362	3014.130127	1322.076904	5752.039551	...	31.121002	41.689007

Fonte: o Autor.

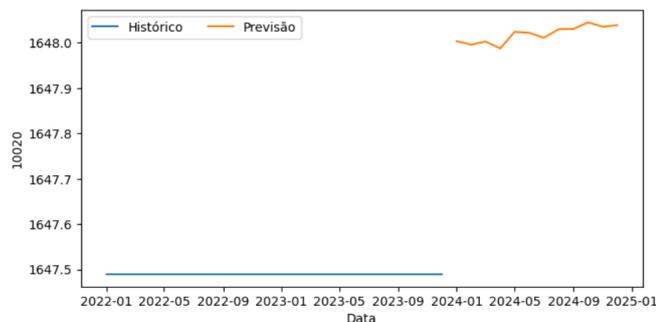
É visualizado a previsão de preço futuro para os materiais, a **FIGURA 14** que possui o menor erro médio apresentando a base inicial de preço em R\$185,00 e o preço futuro de R\$ 179,00, na **FIGURA 15** é apresentado o menor desvio padrão e na **FIGURA 16** é apresentado o maior erro médio sendo os dois devido o tratamento adotado para os dados faltantes.

Figura 14 – Preço futuros (26)



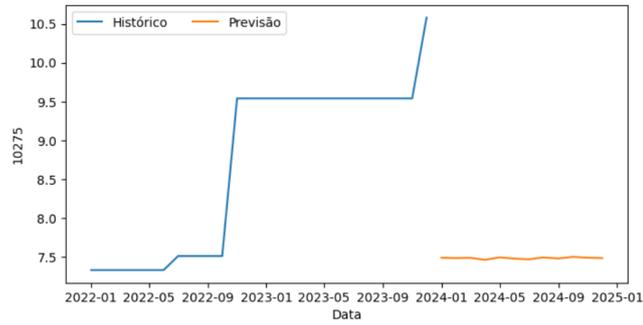
Fonte: o Autor.

Figura 15 – Preço futuros (12)



Fonte: o Autor.

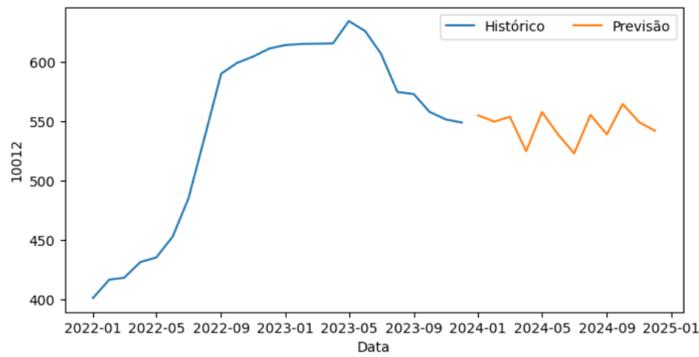
Figura 16 – Preço futuros (56)



Fonte:o Autor.

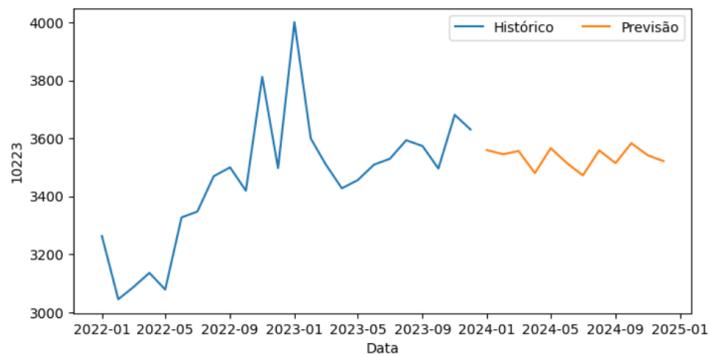
Nas **FIGURAS 17, 18 e 19**, é visualizado a geração do preço futuro para outros materiais selecionados aleatoriamente.

Figura 17 – Preço futuros (10012)



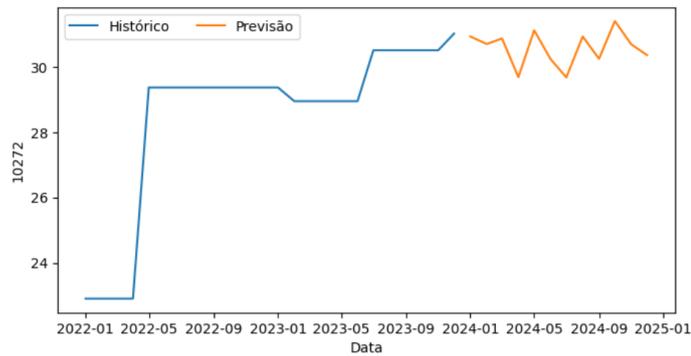
Fonte:o Autor.

Figura 18 – Preço futuros (10223)



Fonte:O Autor.

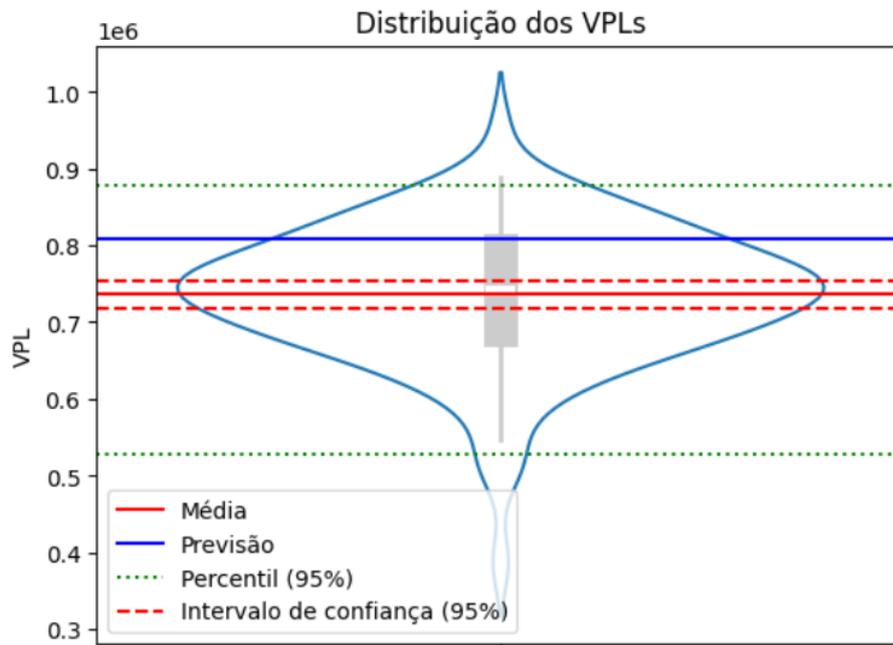
Figura 19 – Preço futuros (10272)



Fonte:o Autor.

A **FIGURA 20**, é apresentado a distribuição dos Valores Presentes Líquidos obtidos através de 100 iterações. A forma de violino utilizada para representar a distribuição permite visualizar a densidade dos dados em diferentes faixas de valores, além de destacar a mediana e os quartis, as linhas horizontais representam a média, a previsão, o percentil 95% e o intervalo de confiança de 95%.

Figura 20 – Gráfico Violino



Fonte:o Autor.

Em resumo, o projeto indica que os Valores Presentes Líquidos gerados têm probabilidade de alcançar o real. No entanto, o valor previsto calculado no sistema de orçamento em evidência na cor azul está distante da média e fora do intervalo de confiança do valor previsto pelo modelo de simulação, isto pode estar ocorrendo devido a série não estar correlacionada e gerando uma variabilidade maior durante a realização da simulação. A utilização de serie correlacionada pode ser trabalhada em trabalhos futuros durante a execução da rede LSTM para análise dos resultados, outro item que pode estar causando é a variação da taxa de desconto utilizada que deve ser avaliada para tomada de decisão final.

Média dos Valores Presentes Líquidos é obtida foi: R\$ 735.849,78, o desvio padrão dos Valores Presentes Líquidos: R\$ 87.899,88 e ele apresenta probabilidade de Valor Presente Líquido negativo: 0,00%

## 5 Considerações Finais

O algoritmo desenvolvido utilizou redes neurais LSTM para prever os preços futuros de diferentes materiais com base em dados históricos e a arquitetura utilizada foi ajustada com a ajuda do *Keras Tuner* para otimização dos hiper parâmetros. Após a previsão dos preços foi possível analisar os erros gerados e calcular a média e o desvio padrão para cada material, diante destes valores os preços futuros previstos foram ajustados considerando os erros médios para elaboração novas previsões mais realistas. Os preços futuros foram utilizados para calcular o custo de produção e integrados com a receita para analisar o impacto financeiro das previsões ajustadas, estas previsões foram utilizadas para calcular o Valor Presente Líquido de diferentes cenários considerando a incerteza das previsões.

Sugere-se para trabalhos futuros a utilização de um banco de dados maior, utilizando as série de dados correlacionadas para redução da variabilidade da simulação e destaca-se ainda a utilização do modelo para detecção de alerta precoce de flutuação de preço e otimização do processo de compra de insumos, considerando tendências de mercado e outros fatores relevantes, que podem influenciar diretamente nas negociações sobre períodos de preço mais baixos a exemplo negociação fornecimento a preço fixo, temos também a possibilidade de análise referente a flutuações cambiais nos custos dos insumos importados que podem ser abordadas.

Em suma, a otimização da previsão de preços de insumos por meio do uso de redes LSTM oferece benefícios para a indústria de celulose, pois o problema proposto foi explorado e tornou visível a utilização de modelos de redes neurais para simulação de custos tornando a tomada de decisão baseada em dados.

# Referências

BAZERMAN M. H.; MOORE, D. A. **Judgment in Managerial Decision Making**. [S.l.]: Wiley, 2017. ISSN 978-1119427384.

CAMPOS, E. S. Curso básico de fabricação de papel. **ABTCP**, 2008.

CARVALHO, J. N. S. d. Lucratividade e desempenho industrial: uma apreciação empírica de duas abordagens teóricas. v. 113, 2006.

CAUCHICK-MIGUEL, P. A.; FLEURY, A.; MELLO, C. H. P.; NAKANO, D. N. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. 3. ed. Rio de Janeiro: Elsevier, 2018. 244 p. ISBN 9788535291346. Disponível em: <<https://integrada.minhabiblioteca.com.br/books/9788595153561>>. Acesso em: 8 abr. 2022.

EMBRAPA. Empresa brasileira de pesquisa agropecuária. 2021. Disponível em: <<http://www.embrapa.br>>. Acesso em: 7 jan. 2024.

GENTLE, J. E. **Random number generation and Monte Carlo methods**. New York: Springer, 2003. 55 p.

HEINRICHS, H. S. E. Sharing economy: A potential new pathway to sustainability. *GAIA-Ecological Perspectives for Science and Society*, p. 228 –231, 2013.

IBGE. Produção da extração vegetal e silvicultura. p. 1–9, 2022.

LIU, J.; CHAO, F.; LIN Y.-C. AND LIN, C.-M. S. Stock prices prediction using deep learning models. 2019. ISSN 1909-12227.

MATZ, A.; CURRY, O. J.; FRANK, G. W. **Contabilidade de Custos**. São Paulo: Atlas, 1978. 534:535 p.

MEHTAB, S.; SEN, J.; DUTTA, A. Stock price prediction using machine learning and lstm-based deep learning models. 2020. ISSN 2009-10819.

OLAH, C. Understanding lstm networks. 2015. Disponível em: <<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em: 15 jul. 2024.

PADOVEZE, C. L. Controladoria estratégica e operacional 3. ed. São Paulo: Cengage Learning, 2012.

QUEIROZ, S. C. S. Influência da densidade básica da madeira na qualidade da polpa kraft de clones híbridos de eucalyptus grandis. **Revista Árvore**, Viçosa, v. 28, n. 6, p. 901–909, 2004.

RAUTENBERG SANDRO; CARMO, P. R. V. d. Big data e ciência de dados: complementaridade conceitual no processo de tomada de decisão. v. 13, n. 1, p. 56–67, 2019.

REZENDE J. L. P.; OLIVEIRA, A. D. Análise econômica e social de projetos florestais. **Viçosa**, Viçosa, 2008.

SANTOS, C. R. D. O movimento competitivo de verticalizaÇÃo da cadeia de celulose: Seria o fim dos fabricantes independentes de papel tissue no brasil? **FGV**, São Paulo, 2018.

SANTOS, J. L.; SCHIMIDT, P.; PINHEIRO, P. R. **Fundamentos de Gestão Estratégica de Custos**. São Paulo: Springer, 2006. 12 p.

TURBAN, E.; SHARDA, R.; DELEN, D.; KING, D. **Business Intelligence: A Managerial Approach**. Prentice Hall, 2011. ISBN 9780136100669. Disponível em: <https://books.google.com.br/books?id=IvZORAAACAAJ>.

# ANEXO A – Tabela de erros

Comp_list	Error 1	Error 2	Error 3	Error 4	Error 5	loc	scale
Material 1	0,5745	0,5849	0,5884	0,5944	0,6205	- 0,0473	0,1994
Material 2	0,7359	0,7403	0,7484	0,7671	0,7671	- 0,0268	0,2814
Material 3	- 1,0277	- 1,0317	1,0559	1,4107	1,4886	- 0,0948	0,4042
Material 4	- 0,9875	- 1,0010	- 1,0165	- 1,0405	- 1,0597	- 0,0220	0,4136
Material 5	0,7454	0,7530	0,7659	0,7844	0,7995	0,0305	0,2320
Material 6	- 0,7826	- 0,7859	- 0,8066	- 0,8200	- 0,8275	- 0,0794	0,2770
Material 7	1,5718	1,6014	1,6031	1,6048	1,6080	0,1684	0,6760
Material 8	- 0,5869	- 0,5920	- 0,6429	- 0,6768	- 0,7512	- 0,0807	0,1745
Material 9	- 0,7488	- 0,7548	- 0,7610	- 0,7624	- 0,7682	- 0,0121	0,2028
Material 10	1,1424	1,1434	1,1438	1,1440	1,1441	0,3108	0,5290
Material 11	- 0,8467	- 0,8535	0,8681	- 0,8834	- 0,8877	- 0,1672	0,3766
Material 12	- 0,0007	- 0,0007	- 0,0007	- 0,0007	- 0,0007	- 0,0004	0,0001
Material 13	0,9621	1,0078	1,0110	1,0203	1,0248	- 0,0474	0,4473
Material 14	1,4876	1,5583	1,5753	1,5874	1,6036	0,1665	0,6263
Material 15	- 0,8340	- 0,8538	- 0,8551	- 0,8723	- 0,9137	- 0,0878	0,3111
Material 16	12,4405	12,4429	13,6144	13,6145	13,6168	0,5205	2,3304
Material 17	8,1549	8,1687	13,0114	13,0244	18,8804	0,3470	1,6869
Material 18	0,8465	0,8479	0,8511	0,8523	0,8524	0,0359	0,2626
Material 19	3,1056	3,1174	3,1247	3,1310	3,1465	0,1372	0,5833
Material 20	0,7871	0,7968	0,7973	0,7992	0,8027	0,0996	0,2440
Material 21	- 0,8864	- 0,9246	0,9983	0,9991	1,1741	- 0,1764	0,4889
Material 22	1,0213	1,0246	1,0279	1,0589	1,1167	- 0,0769	0,4185
Material 23	- 0,4432	- 0,4432	- 0,4505	- 0,5584	- 0,6482	- 0,0737	0,1746
Material 24	0,5807	0,5824	0,5903	0,5985	0,6046	0,0301	0,1679
Material 25	- 0,7764	- 0,7814	- 0,8377	- 0,8678	- 0,8927	- 0,1500	0,2734
Material 26	- 3,7962	- 3,8353	- 3,8367	- 3,8523	12,4624	- 0,1899	0,9797
Material 27	- 0,9959	- 0,9988	- 1,3090	- 1,3119	- 1,5908	- 0,1096	0,3618
Material 28	1,1984	1,2373	1,2421	1,3044	1,3381	- 0,0546	0,6019
Material 29	- 0,8141	- 0,8145	- 0,8150	- 0,8171	- 0,8411	- 0,0295	0,3743
Material 30	- 1,1131	- 1,1542	- 1,1790	- 1,1826	- 1,2172	- 0,2103	0,3811
Material 31	2,3015	2,3214	2,3249	2,3273	2,3311	0,1346	0,7130
Material 32	0,7167	0,7369	0,7477	0,7602	0,7710	- 0,0526	0,2384
Material 33	7,4431	7,5057	7,5140	7,5257	7,5535	0,1246	1,9368
Material 34	- 1,6924	- 1,6924	- 1,7207	- 1,7207	- 1,7386	- 0,2850	0,4691
Material 35	0,4831	0,4991	- 0,5261	- 0,5272	- 0,6076	- 0,0672	0,1455
Material 36	0,8727	0,8750	0,8812	0,8814	0,9311	- 0,0062	0,2784
Material 37	- 1,0061	- 1,0199	- 1,0260	2,2020	2,2459	- 0,1961	0,4445
Material 38	- 1,0760	- 1,1381	- 1,1818	- 1,2069	- 1,3637	- 0,1245	0,3757
Material 39	0,8521	0,8768	0,9658	0,9739	1,0004	- 0,0287	0,2745
Material 40	- 0,6657	- 0,6716	- 0,6985	- 0,7799	- 0,7856	- 0,1173	0,3331
Material 41	- 0,0007	- 0,0007	- 0,0007	- 0,0007	- 0,0007	- 0,0004	0,0001
Material 42	0,9777	0,9791	1,0242	1,0289	1,0632	0,0996	0,4848
Material 43	0,9830	0,9922	1,0015	1,0110	1,0342	0,0253	0,4217
Material 44	0,5266	0,5342	0,5393	0,5439	- 0,5513	0,0154	0,2158
Material 45	- 0,8148	- 0,8152	- 0,8540	- 0,8547	- 0,8568	- 0,0497	0,2951
Material 46	0,8827	0,8894	0,9134	0,9243	0,9602	- 0,0565	0,2736
Material 47	0,8325	0,8365	0,8438	0,8564	0,8609	- 0,0613	0,2857
Material 48	0,7546	0,7613	0,7646	0,7694	0,7709	0,0529	0,4556
Material 49	1,6164	1,6180	1,6287	1,6359	1,6406	- 0,0019	0,5322
Material 50	0,7730	0,7739	0,7752	0,7761	0,7769	0,0556	0,3163
Material 51	0,6461	0,6476	- 0,6510	0,6536	- 0,7649	0,0159	0,2560
Material 52	- 0,4194	- 0,4402	- 0,4617	- 0,5161	- 0,5376	- 0,0628	0,1648
Material 53	0,6815	0,6820	0,6825	0,6831	- 0,7929	- 0,0479	0,2616
Material 54	1,1138	1,1140	1,1141	1,1143	1,1149	0,0319	0,4027
Material 55	0,6016	0,6018	0,6446	- 0,6662	- 0,6716	0,0030	0,2099
Material 56	11,8965	11,9043	11,9067	11,9119	11,9119	2,5075	5,0382
Material 57	3,7317	3,7811	3,9476	3,9888	4,1617	0,3952	1,3579
Material 58	- 1,1878	- 1,1888	- 1,1892	- 1,1895	- 1,1897	- 0,2762	0,6271
Material 59	- 0,9589	- 0,9682	- 0,9763	- 0,9868	- 0,9883	- 0,2535	0,4537
Material 60	- 0,0006	- 0,0006	- 0,0006	- 0,0006	- 0,1406	- 0,0008	0,0070
Material 61	0,8953	- 0,9075	0,9127	- 0,9558	- 1,0458	- 0,0367	0,3995
Material 62	1,9607	2,0488	2,1411	2,1760	2,1922	- 0,0777	0,7886