

Monitorização de Produção com Geração de Relatórios e Alertas

Douglas Menegol Folletto

Dissertação apresentada à Escola Superior de Tecnologia e de Gestão de Bragança para
obtenção do Grau de Mestre em Sistemas de Informação.

No âmbito da dupla diplomação com a Universidade Tecnológica Federal do Paraná

Trabalho orientado por:

Prof. Dr. Paulo Jorge Teixeira Matos

Prof. Dr. Alan Gavioli

Bragança

Dezembro/2018

*Dedico esta dissertação de mestrado aos meus pais, **Mauri** e **Beatriz**, que sempre me incentivaram e apoiaram para que eu realiza-se meus sonhos e objetivos.*

Agradecimentos

Agradeço a Deus, por todas as bênçãos, paciência, sabedoria, fé e força durante este período longe de casa e por ter entregue meu destino nas mãos dessas pessoas tão especiais, meus pais.

Aos meus pais **Mauri** e **Beatriz** pelo carinho, incentivo e confiança, sem os quais esta conquista não seria possível. Agradeço por todo o amor, educação e conselhos que me guiam nos desafios diários da vida. Vocês são meu alicerce de sustentação.

Á todos os meus familiares e amigos que contribuíram de forma direta ou indireta para meu sucesso.

Por fim a Universidade Tecnológica Federal do Paraná, ao Instituto Politécnico de Bragança e a todos os professores, que contribuíram para meu conhecimento e aprendizagem durante meu percurso acadêmico e que colaboraram para o desenvolvimento desta dissertação.

Resumo

A análise de séries temporais estão sempre presentes nos mais diferentes domínios, pela grande facilidade de representar os acontecimentos em relação ao tempo. Alguns dos fatores que podem ser investigados nas séries temporais são a descrição do comportamento da série e/ou a realização de previsões dos valores futuros da série. O principal objetivo desta dissertação foi encontrar um modelo preditivo que conseguisse apresentar valores futuros satisfatórias para as tomadas de decisões empresariais. A metodologia utilizada consistiu de três etapas principais que foram: a definição do problema, seguido da preparação dos dados com uma investigação sobre estes dados para então criar um modelo preditivo e finalmente apresentar os resultados. Identificou-se, na investigação sobre os dados, que grande parte dos produtos não estavam mais sendo produzidos, sendo assim, foi utilizado do k-means para realizar o agrupamento dos produtos com demanda regular e assim prosseguir com a criação do modelo preditivos, em que, onde foram testados os algoritmos de média móvel integrada autoregressiva sazonal (SARIMA), perceptron multicamadas (MLP) e floresta aleatória (RF). Os melhores resultados foram apresentados pelo algoritmo de Floresta Aleatória, com um coeficiente de determinação de aproximadamente 88%, com um erro absoluto médio de aproximadamente 189 unidades de produtos.

Palavras-chave: Séries Temporais, Aprendizado de Máquina, Previsão.

Abstract

The analysis of time series are always present in the most different domains, due to the great ease of representing events in relation to time. Some of the factors that can be investigated in the time series are the description of the behavior of the series and/or the realization of forecasts of future values of the series. The main objective of this dissertation was to find a predictive model that could present satisfactory future values for business decision making. The methodology used consisted of three main steps that were: the definition of the problem, followed by the preparation of the data with an investigation on these data to then create a predictive model and finally present the results. It was identified in the research on the data that most of the products were no longer being produced, so that k-means was used to group products with regular demand and thus to proceed with the creation of the predictive models, where the algorithms of integrated autoregressive seasonal moving average (SARIMA), multilayer perceptron (MLP) and random forest (RF) were tested. The best results were presented by the Random Forest algorithm, with a determination coefficient of approximately 88%, with an mean absolute error of approximately 189 product units.

Keywords: Time Series, Machine Learning, Forecasting.

Conteúdo

1	Introdução	1
1.1	Enquadramento	3
1.2	Objetivos	4
1.2.1	Objetivos Específicos	4
1.3	Motivação	4
1.4	Limitações	7
1.5	Estrutura do Documento	7
2	Técnicas para análise de séries temporais	9
2.1	Série Temporal	9
2.2	Modelos Matemáticos e Estatísticos	11
2.2.1	Suavização Exponencial	11
2.2.2	Modelo Auto-Regressivo Integrado de Médias Móveis	12
2.3	Aprendizado de Máquina	13
2.3.1	Floresta Aleatória	14
2.3.2	Multilayer Perceptron	14
2.3.3	Long short-term memory	15
3	Materiais e Métodos	17
3.1	Etapa 01: Definição do Problema	18
3.2	Etapa 02: Preparação dos dados	18
3.2.1	Feature engineering	19

3.3	Etapa 03: Exploração dos dados	20
3.3.1	Agrupamento da Demanda	20
3.4	Etapa 04: Modelagem Preditiva	20
3.4.1	Mean Absolute Error	21
3.4.2	<i>Mean Squared Error</i> (MSE) e <i>Root Mean Squared Error</i> (RMSE)	22
3.4.3	<i>R-Squader</i> (R^2)	22
3.5	Etapa 05: Visualização dos resultados	22
3.6	Materiais	23
4	Análise Exploratória e Preparação dos Dados	25
4.1	Análise Estatística do Dataset	25
4.2	Agrupamento dos dados	28
5	Resultados e Discussões	31
5.1	Conjunto de Treinamento - Tipo 01	31
5.2	Conjunto de Treinamento - Tipo 02	32
5.3	Conjunto de Treinamento - Tipo 03	33
5.4	Conjunto de Treinamento - Tipo 04	34
6	Aplicação	37
6.1	Descrição dos Processos de Pré-Processamento	37
6.1.1	Carregar arquivos	38
6.1.2	Selecionar variáveis necessárias	38
6.1.3	Realizar o agrupamento dos produtos e selecionar os produtos com previsão estimável	39
6.2	Descrições sobre os Processos do Treinamento do Modelo	40
6.2.1	Separação das variáveis independentes e dependentes e modelação do conjunto de dados nos sub-conjuntos de treinamento e teste	41
6.2.2	Realizar o treinamento do modelo escolhido	42
6.3	Descrições sobre os Processos de Previsão do Modelo	43

6.3.1	Carregar o modelo e as transformação dos dados	43
6.3.2	Variável informada pelo usuário	43
6.3.3	Realizar a previsão com o modelo	44
6.3.4	Exibição dos gráficos e salvamento do arquivo dos dados da previsão	45
7	Conclusões e Trabalhos Futuros	47
7.1	Conclusão	47
7.2	Trabalhos Futuros	48
A	Escola de Verão & Simpósio de Dupla Diplomação 2018	A1
B	9º MEDITEC – Medianeira in Technology	B1
C	Tabela dos resultados da previsão realizados pelo modelo RF entre as semanas 5 a 9 do ano de 2018	C1

Lista de Tabelas

4.1	Amostra dos dados originais fornecidos pela fabricante	26
4.2	Amostra dos dados de vendas e depois de devolução	27
4.3	Amostra dos dados de vendas diárias	27
4.4	Comparação entre as classificações dos produtos	30
5.1	Resultados das métricas de avaliação para o modelo de <i>Random Forest</i> (RF) . .	32
5.2	Resultados das métricas de avaliação para o modelo de SARIMAX	33
5.3	Resultados das métricas de avaliação para o modelo de RF	34
5.4	Amostra do conjunto de treinamento do tipo 04	34
5.5	Resultados das métricas de avaliação para o modelo de Multilayer Perceptron (MLP) e LSTM	35
C.1	Resultados da previsão realizado pelo modelo da aplicação para a semana do ano 5 a 9 do ano de 2018	C2

Lista de Figuras

3.1	Etapas da metodologia de trabalho	17
3.2	Categorização da demanda por meio dos valores do <i>Average Demand Interval</i> (ADI) e do Coeficiente de Variação ao Quadrado (CV^2), adaptado de [27]	21
4.1	Sazonalidade dos dados agrupados da série temporal	29
5.1	Importância dos atributos para o algoritmo floresta aleatória	32
5.2	Comparação entre o y_{real} e o $y_{previsto}$ por meio do modelo RF	32
5.3	Comparação entre o y_{real} e o $y_{previsto}$ obtido por meio do modelo SARIMAX	33
6.1	Diagrama com as atividades executadas no pré-processamento dos arquivos de dados brutos	38
6.2	Tela do gerenciamento dos arquivos brutos do conjunto de dados	40
6.3	Diagrama com as atividades executadas no treinamento do modelo	41
6.4	Tela das tarefas de previsão de treinamento do modelo	42
6.5	Diagrama com as atividades executadas na previsão com o modelo	43
6.6	Dashboard dos produtos previstos com os gráficos e filtros	46

Acrônimos e Siglas

CV² Coeficiente de Variação ao Quadrado. xv, 20, 21

R² *R-Squader*. xii, 22, 31, 33, 34, 48

ADI *Average Demand Interval*. xv, 20, 21

ARIMA Modelo Auto-Regressivo Integrado de Médias Móveis. 3, 12

ARMA Auto-Regressivos de Médias Móveis. 3, 12

CEC *Constant Error Carousel*. 15

ERP Sistema Integrado de Gestão Empresarial. 48

IDE Ambiente de Desenvolvimento Integrado. 37

JIT *Just in Time*. 5, 6

logistic Função Logística ou Sigmoide. 15

LSTM *Long short-term memory*. 3, 15, 34

MAE *Mean Absolute Error*. 21, 22, 34

ML *Machine Learning*. 3, 6, 7, 13, 18, 47

MLP Multilayer Perceptron. xiv, 14, 15, 34, 35, 48

MSE *Mean Squared Error*. xii, 22

PDP *Production Manager Program*. 3, 18

relu Função de Ativação Retificada. 15

RF *Random Forest*. xiv, 3, 14, 31, 32, 34, 47, 48

RMSE *Root Mean Squared Error*. xii, 22, 34

RNN *Recurrent Neural Network*. 15, 16

SARIMAX Média Móvel Integrada Auto-regressiva Sazonal. 31

tanh Função Tangente Hiperbólica. 15

Capítulo 1

Introdução

As organizações proporcionaram um grande avanço no mundo moderno, tornando-se impossível imaginar um mundo sem as mesmas. Conforme as definições apresentadas em [1] as organizações são formadas por duas ou mais pessoas que compartilham de um objetivo ou um conjunto de objetivos em comum e trabalhando juntas para alcançá-los.

As organizações são classificadas em três setores fundamentais:

- setor primário: são as organizações que estão relacionadas à exploração dos recursos naturais.
- setor secundário: são as organizações responsáveis pela produção (fabricação ou montagem) de alguns produtos.
- setor terciário: são as organizações que prestam algum tipo de serviço para as empresas do setores primário, secundário ou para o consumidor.

A estruturação dentro das organizações acontecem por meio da administração. O modelo administrativo utilizado se relaciona diretamente com o sucesso ou fracasso de qualquer entidade [1]. Está estruturação é realizada por meio de processos que definem como serão realizadas as atividades nas organizações ou como decorrerá a utilização dos recursos, proporcionando assim, alcançar os objetivos/metasp da organização.

O processo é a forma sistemática de fazer alguma coisa. Todos os processos ocorrem por meio de ciclos, sendo alguns mais utilizados que outros. O planejamento é necessário para qualquer processo, independentemente do grau de importância e abrangência que o mesmo terá na organização [1].

Todas as organizações tem pelo menos cinco atividades básicas: atividades de marketing/vendas, contábeis, gestão de recursos-humanos, logística e operações. Cada um com seu grau de importância. As atividades de marketing/vendas estão relacionadas com a procura de demanda, e possuem como atuadores as áreas de marketing e de vendas. As atividades contábeis desenvolvem o papel de fornecer informações úteis aos usuários da contabilidade para a tomada de decisões, tais como a contabilidade financeira, de custos e gerencial. As atividades de gestão de pessoas estão relacionadas ao controle e gerenciamento dos recursos humanos e da organização, afim de conciliar os objetivos de ambas as partes. As atividades logísticas são associadas à gestão dos materiais físicos precisos para o funcionamento da organização. Por fim, as atividades de operação são realizadas no processo produtivo, ou seja, na produção dos bens ou serviços que serão disponibilizados pela organização [1].

Todas as atividades são importantes dentro da organização, neste trabalho, abordaremos um processo que acontece dentro das atividades marketing/vendas, previsão de demanda, que implicam diretamente nas atividades logísticas e operacionais.

Em 1980, Jaynes questiona-se, com a informação disponível, quais serão as melhores previsões obtidas sobre os fenômenos observáveis [2]. Assim a análise preditiva é utilizada para a identificação de padrões nos dados e então indicar tendências, mudanças, hábitos de consumo, entre outras coisas mais.

Esta pergunta ecoa até hoje em busca de soluções. Isso acontece por causa de existirem diferentes casos, assim, são criadas soluções generalizadas para resolver este problema.

Desde 1980 até os dias atuais, ocorreu uma evolução muito grande do poder computacional e a quantidade de dados aumentou exponencialmente nos últimos anos. Com isso aumentou a motivação de extrair informações importantes, que podem agregar valor nas tomadas de decisões empresarias.

A previsão de vendas é uma estimativa de quanto será vendido por uma empresa num determinado período de tempo. Assim, a contribuição da previsão de vendas no processo do planejamento corporativo é amplamente necessário e reconhecido [3]. Uma das áreas em que a previsão de vendas pode ser aplicado é na elaboração do *Production Manager Program* (PDP), que é a elaboração o plano de produção dos produtos que serão manufaturados pela linha de produção.

Uma previsão de demanda de alta qualidade, permite que a empresa forneça aos seus clientes os produtos e serviços, quando e onde eles desejarem [4]. Com isso, a procura para melhorar os processos de previsão de demanda tornou-se necessário, devido ao auxílio fornecido no planejamento da cadeia de suprimentos, fornecimento futuro, estratégias de vendas e assim por diante.

1.1 Enquadramento

A utilização dos modelos de séries temporais em previsão de demanda, procura encontrar um modelo estatístico que descreva os acontecimentos passados para realizar as projeções de demandas futuras [1]. As técnicas utilizadas para a previsão de séries temporais podem ser divididas em dois grupos: técnicas tradicionais baseadas em modelos matemáticos e estatísticos, como Auto-Regressivos de Médias Móveis (ARMA) e Modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA), e as técnicas de *Machine Learning* (ML), como *Long short-term memory* (LSTM) e RF.

Uma comparação entre os métodos de previsão utilizando 1045 séries temporais é apresentado em [5], em que se concluiu que as técnicas de ML apresentam resultados inferiores as técnicas estatísticas tradicionais. Constatou-se que os artigos publicados, que utilizam técnicas de ML, não realizam comparação com os modelos matemáticos e estatísticos tradicionais, assim, não existem provas experimentais que as técnicas de ML são mais precisas. Um fator que deve ser levado em consideração é a dificuldade de modelagem de algumas técnicas de ML, além do grande poder computacional que é necessário para o treinamento dos modelos de previsão.

Assim sendo, nota-se a importância de realizar trabalhos que abordem também as técnicas de previsão tradicionais, realizando a comparação entre elas, afim de encontrar a técnica com o melhor desempenho.

1.2 Objetivos

Realização da previsão de demanda para os produtos manufaturados de uma indústria automobilística, diminuindo o erro das previsões com uma antecipação mínima de 5 semanas, gerando assim o plano de produção.

1.2.1 Objetivos Específicos

- Realizar a previsão de vendas com uma antecipação mínima de 5 semanas;
- Identificação dos produtos manufaturados que possuem uma regularidade e serão utilizados na solução desenvolvida;
- Comparação das técnicas utilizadas na previsão;
- Visualização do plano de produção para as semanas previstas, podendo utilizar os resultados em outro sistema desenvolvido pela Techwelf ¹.

1.3 Motivação

A dificuldade em encontrar o equilíbrio entre a necessidade de reduzir os custos e a necessidade de melhorar o nível de serviço, acaba por favorecer uma das opções. Por exemplo, ao reduzir o stock, em busca de redução dos custos, prejudicaria o rápido atendimento na variação de demanda do cliente. Por outro lado, o aumento no stock de segurança, para atender a demanda durante os picos, gera um excesso de stock e um aumento nos custos [6].

¹<http://www.techwelf.com/>

As principais estratégias dos processos em uma cadeia de suprimento são divididas em duas categorias: na antecipação aos pedidos do clientes, produção puxada, ou na resposta ao pedidos dos clientes, produção empurrada [7].

Na produção puxada, *pull*, o processo inicia-se em resposta ao pedido do cliente, com isso, é produzido o que realmente o cliente precisa. Os processos *pull* podem ser definidos como processos reativos, porque reagem à demanda do cliente [7].

As principais vantagens da produção puxada é o *lead time* de pedido em um intervalo menor, isso colabora com um nível de stock menor. Reação mais rápida às oscilações e incertezas da demanda [8].

Na produção empurrada, *push*, o processo inicia-se na antecipação aos pedidos dos clientes. A quantidade que será produzida é calculada por meio de previsões, podendo não ser a real necessidade do cliente. Este processo pode ser chamado de especulativo, porque responde as previsões e não a demanda real do cliente [7].

As principais desvantagens dos processos empurrados são [8]:

- O tempo de armazenagem do stock nos depósitos, que podem ser de longos períodos, aumentando o risco do produto ficar obsoleto, custos de manutenção e diminuição de lucros;
- A acumulação de stocks pode limitar a capacidade da empresa acompanhar as tendências do mercado.

A aplicação dos dois processos *push/pull* juntos na cadeia de suprimento torna-se muito viável. A adaptação dos dois processos na cadeia de suprimentos depende de uma análise, em que os gestores definiram as melhores abordagens para sua organização.

Dentro desses conceitos, surge o sistema *Just in Time* (JIT), produção puxada, foi criado por Taiichi Ohno, na Toyota Motor Corporation, com a proposta de produzir e atender no tempo certo à produção, ou seja, disponibilizar o produto no lugar certo e na hora certa [9].

Com o objetivo de reduzir os processos que consomem recursos ou causam desperdício, por exemplo o stock, o transporte interno de mercadoria, paradas durante o processo,

retrabalho, entre outros, deveriam ser suprimidos da linha de produção de fábricas.

Assim o sistema JIT tem como princípios [10]:

- Trabalhar com o menor stock possível ou stock zero;
- Produção enxuta, uma produção sem desperdícios ou com o mínimo de desperdício possível;
- Esforço constante na resolução dos problemas;
- Lotes de fornecimentos reduzidos. Os *lead times* do fornecimento são reduzidos, permitindo oferecer um alto nível de qualidade.
- Redução da quantidade de stock armazenada, devido ao fato de se produzir o necessário, reduzindo os problemas de perda de stock, diminuição nos lucros, produto obsoleto ou/e espaço para o armazenamento.

O JIT proporciona uma redução de custos, de stock e melhoria na qualidade, proporcionando a obtenção de maiores lucros e melhoria no retorno do capital investido.

Um dos problemas que acontecem com a previsão de vendas é a baixa acurácia que provoca problemas na estimativas, requerendo um elevado nível de stock de segurança, e isso não deveria ser uma solução [11]. Devido a essa situação e procurando aplicar as melhorias desenvolvidas nos modelos de previsão apresentados nas pesquisas acadêmicas, surgiu a oportunidade do desenvolvimento desse trabalho buscando a aproximação entre os avanços das pesquisas acadêmicas e a indústria, validando os novos modelos e analisando seu comportamento em uma situação real.

A aplicação de técnicas de inteligência artificial nas empresas está nos assuntos mais desenvolvidos dos últimos tempos, assim a contribuição desse trabalho no meio acadêmico com a aplicação de técnicas de ML em situações reais fornecendo um *insight* que possa ser utilizado nas decisões empresariais.

Outro ponto a se destacar é na utilização das previsões no nivelamento de produção, que significa a divisão do ciclo mensal em quantidades diárias [12], assim a linha de produção atua com ciclos diários na manufatura dos produtos.

O processo de nivelamento é composto por várias operações de planejamento em busca da distribuição das ordens de lotes programados em sequências otimizadas para a produção. Essa sequência deve respeitar a capacidade produtiva e suavizar as quantidades que serão produzidas [13].

O processo pode ser descrito da seguinte forma:

- Dimensionamento dos lotes em quantidades menores;
- Nivelamento da carga de produção;
- Escalonamento dos produtos respeitando a capacidade diária;
- Sequência de manufatura para linha de produção.

1.4 Limitações

Uma das limitações do trabalho encontra-se no tamanho do conjunto de dados fornecidos pelo fabricante, que é de apenas dois anos de dados. Conforme Hyndman, Kostenko et al., as séries temporais curtas possuem muitas dificuldades de fornecer uma previsão confiável. Normalmente, os dados reais contém muita variação aleatória, com isso surge a necessidade de utilizar mais amostras [14]. Apesar dos novos pedidos feitos, a empresa em questão não teve como disponibilizar mais dados para análise.

1.5 Estrutura do Documento

O trabalho está organizado da seguinte forma:

No capítulo 2 apresenta uma introdução ao conceito de séries temporais e aborda as técnicas estatísticas e ML utilizadas para realizar previsões em séries temporais.

O capítulo 3, realiza uma descrição dos processos realizados durante o desenvolvimento do trabalho, abordando as ferramentas e bibliotecas utilizadas.

No capítulo 4 é realizada uma apresentação do dataset utilizado neste trabalho e a identificação dos produtos com demanda regular.

No capítulo 5 são apresentados e debatidos os resultados encontrados, com comparações entre as técnicas apresentadas no capítulo 2.

O capítulo 6 contém as informações sobre o desenvolvimento de protótipo para a utilização na empresa, visando a aplicação dos dados obtidos.

No capítulo 7 são apresentadas as conclusões do trabalho, com comentário sobre os resultados obtidos e sugestões para trabalhos futuros.

Capítulo 2

Técnicas para análise de séries temporais

2.1 Série Temporal

Um conjunto de observações que estão ordenadas no tempo é uma característica de uma série temporal [15]. Os conjuntos de observações pode ser preços de ações [16], sinais de áudio [17] ou imagem por ressonância magnética funcional [18], são alguns exemplos de séries temporais. A variável tempo está num intervalo de segundos, dias, semanas, meses ou anos.

O conjunto de observação pode ser composto de variáveis quantitativas, que apresentam medidas em escala quantitativa, valores numéricos. Essas podem ser contínuas, que assumem valores em uma escala contínua, ou variáveis discretas, que são números finitos ou infinitos contáveis.

Existem dois tipos de séries temporais: estacionárias e não estacionárias. Uma série temporal estocástica é estacionária se a média, variância e autocovariância, em diferentes períodos, forem constantes. Essa característica é encontrada em diferentes períodos, ou seja, não variam com o tempo. Já um série temporal estocástica não-estacionária é quando a média ou a variância variam com o tempo, com isso, somente alguns períodos são

semelhantes, isto prejudica as previsões destas séries temporais.

Os principais fatores a serem investigados em séries temporais são, a simulação com os dados observados para gerar novos valores para o mesmo período ou em cenários futuros. Com a análise estatística de gráficos ou descritiva, procura-se descobrir o comportamento dos dados. Por fim, com a previsão espera-se criar um modelo estatístico utilizando os dados e assim realizar estimativas para os próximos valores das séries, chamado de horizonte de previsão.

Os principais componentes das séries temporais são:

- Tendência (T_t): Representa as alterações unidirecionais ao longo do tempo. Esse tipo de comportamento está presente quando a série exibe crescimento ou decaimento constantes, em períodos consecutivos de tempo.
- Sazonalidade (S_t): O movimento de oscilação ao longo do tempo, com acontecimento em períodos conhecidos ou fixos, por exemplo, os feriados, datas comemorativas do país, entre outros;
- Ciclo (E_t): Representa a variação aleatória em torno dos componentes anteriores.

A decomposição de uma série temporal pode ser representada em dois modelos: aditiva, Equação 2.1, em que as componentes são somadas, ou multiplicativa, Equação 2.2, no qual as componentes são multiplicadas.

$$\textit{Aditivo} \quad Y_t = T_t + S_t + E_t \quad (2.1)$$

$$\textit{Multiplicativo} \quad Y_t = T_t * S_t * E_t \quad (2.2)$$

No processo de análise de série temporal encontra-se o teste de normalidade em que se verifica se os dados são normalmente distribuídos, um teste aplicado é o de Shapiro-Wilk. No teste de raiz unitária é verificado se as propriedades da série não se alteram no tempo, ou seja, se ela é estacionária, sendo um teste aplicado o Dickley-Fuller.

2.2 Modelos Matemáticos e Estatísticos

Esses modelos pressupõe que as observação passadas possuem as informações sobre o comportamento da série temporal, em que o padrão é periódico no tempo. Com a proposta de identificar esse padrão, abstraindo os ruídos que possam estar contidos nas observações e ao identificar o padrão da série é possível realizar previsões para períodos futuros.

2.2.1 Suavização Exponencial

A suavização exponencial é um dos modelos utilizados nas séries temporais, devido a sua simplicidade, facilidade de ajuste e boa precisão. Para efetuar a previsão, é realizada a soma ponderada das observações passadas, com um peso exponencial decrescente nas observações anteriores.

A Suavização Exponencial Simples, utiliza os dados univariados sem tendência ou sazonalidade. Assim esse método requer somente um coeficiente de suavização, chamado de alfa α , que controla a taxa de influencia da observação no tempo [19]. A fórmula de previsão é apresentada na Equação 2.3, no qual o intervalo varia de $0 < \alpha < 1$.

$$\hat{y} = \sum_{j=1}^t \alpha(1 - \alpha)^{t-j} y_j + \alpha(1 - \alpha)^t l_0 \quad (2.3)$$

Na Suavização Exponencial Dupla, Equação 2.4, utiliza de base a Suavização Exponencial Simples com a adição da tendência na série temporal. O coeficiente de suavização beta, β , controla o decaimento da Influência da mudança da tendência, como apresentado na Equação 2.6. O modelo suporta as tendências de diferentes maneiras: aditivo e multiplicativo.

Para as previsões multi-passos a tendência pode ter um comportamento irreal, assim para amortecer a tendência a longo prazo é adicionado o coeficiente de amortecimento

phi, (ϕ), que controla a taxa de amortecimento.

$$\text{Previsão} \quad \hat{y} = l_t + hb_t \quad (2.4)$$

$$\text{Cíclico} \quad l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (2.5)$$

$$\text{Tendência} \quad b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \quad (2.6)$$

Por fim, a suavização exponencial tripla, é a última extensão da suavização exponencial, com a adição da sazonalidade na série temporal univariada. Além dos coeficientes comentados anteriormente, tem a adição do coeficiente gamma, γ , que controla a influência da componente sazonal. Essa é a variação mais avançada da suavização exponencial. Para garantir a sazonalidade é necessário indicar quando o período sazonal acontece.

$$\hat{y} = l_t + hb_t + s_{t-m+h_m^+} \quad (2.7)$$

$$h_m^+ = \lfloor (h - 1) |m| \rfloor + 1 \quad (2.8)$$

2.2.2 Modelo Auto-Regressivo Integrado de Médias Móveis

Este modelo Auto-Regressivos de Médias Móveis (ARMA) é apresentado na Equação 2.9, em que a previsão inclui os valores anteriores dos erros de y_t e os atrasos. Também nota-se que os coeficientes garantem a estacionariedade e invertibilidade da série analisada [5].

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \phi_1 e_{t-1} + \dots + \phi_q y_{t-q} + e_t \quad (2.9)$$

Da combinação da equação 2.9 com a diferenciação surge o Modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA), com as seguintes componentes $ARIMA(p, d, q)$, onde o termo AR é o percentual de diferença, MA é o termo de média móvel e I é a diferenciação, assim é possível tornar uma série não-estacionária em estacionária por

diferenciação:

- p = ordem de termos auto-regressivos;
- d = grau da primeira diferenciação;
- q = ordem de termos de média móvel-

2.3 Aprendizado de Máquina

Devido à grande quantidade de dados disponíveis e à limitação humana para compreender e analisar essas informações, surge o *Machine Learning* (ML). Ele é utilizado para realizar o processamento em grande escala e para conjuntos de dados complexos com grande número de variáveis e recursos. Um dos pontos a se destacar no ML e *Deep Learning*, é o melhor desempenho em grande conjunto de dados melhorando assim, os resultados dos modelos analíticos e preditivos das séries temporais [20].

O ML refere-se à extrapolação dos padrões mapeados nos conjuntos de dados ou na capacidade de realizar previsões de novos dados com o que foi aprendido com os dados conhecidos anteriormente. As duas grandes divisões no aprendizado de máquina são o aprendizado supervisionado e o não-supervisionado, existindo uma terceira conhecida por aprendizado por reforço.

No aprendizado supervisionado, os algoritmos usam os conjuntos de dados categorizados para então classificar os dados semelhantes não categorizados. A categorização dos dados pode ser discreta ou contínua.

No aprendizado não supervisionado não existe uma categorização, o algoritmo aprende sozinho a classificar. Um exemplo é a técnica de agrupamento, em que se procura separar os dados em subconjuntos. Uma das características para os *cluster* são ter grande semelhança entre os dados da mesma classe e baixa similaridade com as demais classes. Existem muitas técnicas utilizadas no aprendizado não supervisionado e o grande sucesso da *Deep Learning* está relacionado a essa eficiência [20].

Por fim, o aprendizado por reforço é semelhante ao aprendizado supervisionado, em que é fornecidos *feedback* para melhorar o desempenho. Uma diferença entre esse aprendizado é que não existe uma categorização nos dados, mas uma avaliação por métricas, por exemplo [20].

2.3.1 Floresta Aleatória

A floresta aleatória, conhecido também com RF é um algoritmo flexível e fácil de ser usado. Além de produzir resultados excelentes na maioria das vezes, mesmo sem o ajuste de hiper-parâmetros [21].

Uma árvore de decisão simboliza uma função que tem como entrada um vetor de valores de atributos e retorna uma decisão [22]. É realizado uma combinação, *ensemble*, de várias árvores de decisão, que tem a finalidade de realizar o aumento do resultado geral, nos critérios de acurácia e estabilidade.

A avaliação da importância do atributo é uma parte importante do processo, pois podemos identificar os atributos mais importantes para a decisão nas RF e remover os atributos menos importantes, ganhando desempenho [21].

2.3.2 Multilayer Perceptron

O algoritmo de aprendizado supervisionado, MLP, tem por definição a função de aprendizado $f(.) : R^m \rightarrow R^0$, em que m é o número de dimensões de entrada e 0 é o número de dimensões de saídas. Dado um conjunto de dados $X = x_1, x_2, \dots, x_m$ e um *target* y , o algoritmo pode aprender uma função de aproximação não linear para classificação ou regressão. Entre a camada de entrada e a camada de saída, podem existir mais camadas, conhecidas como camadas ocultas [23].

A camada que fica à esquerda é conhecida como camada de entrada, que possui um conjunto de neurônios de entrada, que são representados por $x_i | x_1, x_2, \dots, x_m$. Os neurônios da(s) camada(s) oculta(s) transformam os valores da camada anterior em um somatório linear ponderado $w_1x_1, w_2x_2, \dots, w_mx_m$, seguidos por um função de ativação não linear

$g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$, sendo elas:

- Função Identidade: $f(x) = x$
- Função de Ativação Retificada (relu): $f(x) = \max(0, x)$
- Função Tangente Hiperbólica (tanh): $f(x) = \tanh(x)$
- Função Logística ou Sigmoidal (logistic): $f(x) = \frac{1}{(1+\exp(-x))}$

Dentre as vantagens da MLP são a capacidade de aprender modelos não lineares e aprender modelos em tempo real. Nas desvantagens da MLP são que as camadas ocultas possuem funções de perdas não convexas, em que existe mais de um mínimo local.

O treinamento da rede MLP é realizado pela técnica *backpropagation* sem nenhuma função de ativação na camada de saída. A função de perda utilizada é o erro quadrático, Equação 2.10, em que a iniciação dos pesos aleatórios é minimizada pela função de perda, realizando a atualização desses pesos.

$$Perda(\hat{y}, y, W) = \frac{1}{2} \|\hat{y} - y\|_2^2 + \frac{\alpha}{2} \|W\|_2^2 \quad (2.10)$$

Devido a sensibilidade no dimensionamento de recursos, é recomendado a transformação que padronize os dados para ter a média 0 e a variância 1.

2.3.3 Long short-term memory

A arquitetura LSTM é um exemplo de *Recurrent Neural Network* (RNN). Essa arquitetura é conhecida por preservar as informações por longos períodos, sendo um comportamento comum e fácil de aprender. Introduzidas por Hochreiter e Schmidhuber, em 1997, com o objetivo de melhorar o desempenho ao resolver o problema do desaparecimento de gradiente que as RNN naturalmente sofrem ao lidar com grandes quantidade de dados [24].

A célula de memória das redes LSTM possui quatro componentes: Uma célula *Constant Error Carousel* (CEC), que é responsável por percorrer toda a cadeia, não possuindo

uma função de ativação, com isso o gradiente não desaparece no treinamento do *backpropagation*. Isso solucionou o problema das RNN e facilitou a aprendizagem a longo prazo, pois as informações circulam facilmente entre as células inalteradas [25]. Também possui três *gates*, portões, que são formas de, opcionalmente, deixar passar as informações. Ao não deixar passar a informação, expressa o esquecimento. São compostas por uma camada de rede neural sigmoide e uma operação de multiplicação.

Capítulo 3

Materiais e Métodos

Os procedimentos realizados durante o desenvolvimento deste trabalho de dissertação, encontram-se descritos de forma detalhada nesta seção e resumidas na Figura 3.1. A compreensão dos fatos permite a elaboração de previsões e com a ajuda da análise de dados é possível explorar o passado, para a criação de modelos preditivos. Os passos a serem seguidos na análise de dados, conforme apresentado por [26], são: Definição do Problema; Obtenção dos dados; Limpeza dos dados; Normalização dos dados; Transformação dos dados; Estatística exploratória; Visualização exploratória; Modelagem preditiva; Validação do modelo; Visualização e interpretação dos resultados obtidos; e Implementação da solução.

As etapas executadas de forma resumida, são apresentadas no Figura 3.1:

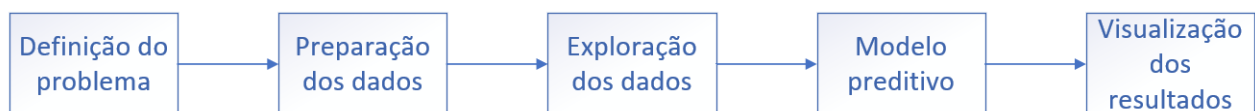


Figura 3.1: Etapas da metodologia de trabalho

3.1 Etapa 01: Definição do Problema

A pergunta base deste trabalho foi: Com base nos dados históricos, é possível realizar uma previsão da quantidade de vendas dos produtos com alta qualidade e precisão para sustentar a definição do *Production Manager Program* (PDP)?

A partir da definição do objetivo, realizou-se a estruturação da metodologia e das ferramentas que se utilizaram.

3.2 Etapa 02: Preparação dos dados

Os passos que foram executados para a obtenção, limpeza, normalização, transformação dos dados em um conjunto ideal, evitando os problemas que os dados brutos possuem valores inválido, ausentes ou outlier¹.

Durante esta etapa, a avaliação da relação entre as variáveis foi realizada e foram removidas as variáveis com muitos valores nulos, selecionado apenas as variáveis que possuíam valores úteis para a análise. A ferramenta utilizada foi a linguagem de programação Python e a biblioteca Pandas, que fornece métodos de leitura de arquivos e manipulação dos dados, assim foi realizado a remoção de colunas e valores ausentes.

O próximo passo foi realizar o agrupamento das vendas por semana do ano, que aconteceu da seguinte forma. Para cada data de entrega foi calculado o primeiro dia daquela semana do ano, após isso foi realizado o agrupamento das quantidades produzidas dos produtos por semana do ano, formando assim três colunas: `Material_ID`, data da semana do ano e Quantidade.

Devido a necessidade de normalização do dados para a aplicação de algumas técnicas de ML, foi utilizados a biblioteca `scikit-learn`² e as transformações `MinMaxScaler` e `StandardScaler`.

A transformação `MinMaxScaler` é utilizada para dimensionar os valores para uma média zero, com escala de variação. A equação utilizada é apresentada em 3.1, em que

¹Dados que se diferenciam drasticamente de todos os outros

²<http://scikit-learn.org/stable/index.html>

p_min e p_max é o intervalo desejado para os dados transformados, por exemplo (0,1) ou (-1,1) e X é o conjunto de dados que será transformado:

$$x_std(X) = \frac{X - min(X)}{max(X) - min(X)} \quad (3.1)$$
$$X_scaled = x_std * (p_max - p_min) + p_min$$

StandardScaler transforma os dados de modo que tenham média zero e desvio padrão igual a um. A Equação 3.2, demonstra como é realizada esta transformação.

$$z = \frac{x - media}{desviopadrao} \quad (3.2)$$

3.2.1 Feature engineering

Um dos objetivos do modelos preditivos é a obtenção dos melhores resultados possíveis. Para alcançar isso, é necessário obter o máximo aproveitamento dos dados que possui. As características dos dados, influenciaram diretamente no desempenho dos modelos preditivos elaborados. *Feature engineering* torna-se uma parte importante no processo, pois com ela, é possível extrair características úteis que podem ser determinantes para o desempenho do modelo.

As tarefas de preparação dos dados e a *feature engineering* podem parecer similares, mas a grande diferença é o tempo para realizar cada tarefa e a complexibilidade. Enquanto a preparação dos dados é um processo mais de limpeza, como a detecção de valores duplicados e valores nulos ou outliers, já a *feature engineering* realiza uma avaliação mais profunda para o conhecimento da variável explorada, por exemplo, redução da dimensionalidade, transformação dos dados ou combinações entre variáveis.

Feature engineering é a de parte de modelar manualmente como deverá ser a entrada, *input*, para algoritmos que serão utilizados na previsão.

3.3 Etapa 03: Exploração dos dados

Nesta etapa acontece a busca dos padrões e relações entre os dados. São utilizados gráficos e a análise estatística ou outros métodos descritos na literatura.

3.3.1 Agrupamento da Demanda

O processo do agrupamento da demanda auxilia na escolha dos métodos de previsão. Desenvolvida por [27] em 2005, os itens são agrupados em quatro categorias: (a) Errático: são os itens com alta variação da quantidade e com períodos regulares; (b) Irregular: são os produtos com alta variação da quantidade e com muitos períodos sem demanda; (c) Regular: são os itens com baixa variação da quantidade e com pedidos regulares e (d) Intermitente: são os produtos com baixa variação na quantidade e com períodos sem demanda.

Os cálculos do ADI, equação 3.3, e do CV^2 , equação 3.4, são apresentados a seguir:

$$ADI = \frac{\text{Número de Períodos Com Demanda}}{\text{Número de Períodos Sem Demanda}} \quad (3.3)$$

$$CV^2 = \left(\frac{\text{Desvio Padrão}}{\text{Média}} \right)^2 \quad (3.4)$$

A Figura 3.2 apresenta os intervalos dos valores entre do ADI e do CV^2 em que é identificada a categoria à qual o item pertence.

3.4 Etapa 04: Modelagem Preditiva

Neste processo, foi elaborado a criação e escolha do modelo que será utilizado para prever os resultados.

Os modelos apresentados nas seções 2.2 e 2.3 foram implementados na linguagem de programação Python por meio das bibliotecas Keras³ e scikit-learn.

³<https://keras.io/>

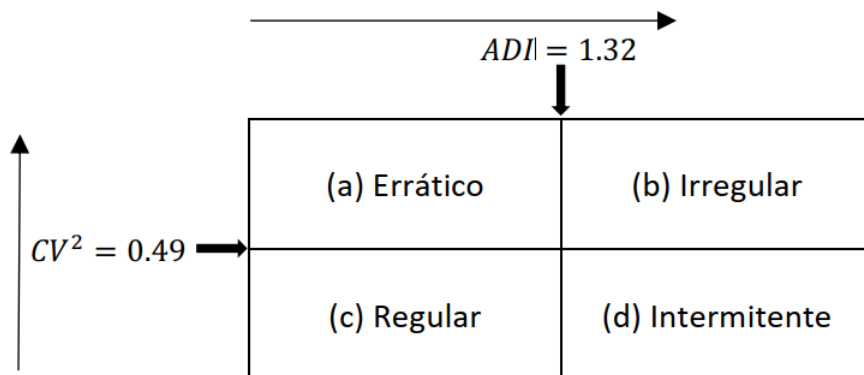


Figura 3.2: Categorização da demanda por meio dos valores do ADI e do CV^2 , adaptado de [27]

Primeiramente, os conjuntos de dados criados foram separados em dois sub-conjuntos *Train*, utilizado para o treinamento do modelo, e *Test*, utilizado para a validação deste modelo.

As amostras foram separadas de forma aleatória por meio do método *train_test_split* da biblioteca scikit-learn. O conjunto original foi dividido em 90% para o sub-conjunto *Train* e 10% para o sub-conjunto *Test*. No sub-conjunto de *Train* foram aplicadas as técnicas que realizando os cálculos das equações e parâmetros para criaram o modelo. O sub-conjunto de *Test* foi utilizado para verificar da precisão do modelo.

Para comprovar a eficiência do modelo de regressão são utilizadas as funções de perda, de pontuação e de utilidade. Isto fornece uma forma de avaliação dos modelos desenvolvidos. As métricas de avaliações que foram utilizadas para medir o desempenho do modelo desenvolvido, estão descritas a seguir:

3.4.1 Mean Absolute Error

O Erro Absoluto Médio, *Mean Absolute Error* (MAE), é uma medida utilizada para avaliar os modelos de regressão. Ela é uma medida de quanto o valor previsto desviou do valor real, em que, quanto maior o valor da medida, maior é o desvio do valor real.

A equação para calcular os valores absolutos dos erros é apresentada em 3.5, onde y

são os valores verdadeiros, \hat{y} são os valores previstos e n é o número de amostras.

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} |y_i - \hat{y}_i| \quad (3.5)$$

3.4.2 MSE e RMSE

O Erro Quadrático Médio é semelhante ao MAE, só que em vez de assumir os valores absolutos dos erros, a diferença entre o valor real e o valor previsto é elevada ao quadrado. Isso torna a métrica sensível aos valores discrepantes, devido à penalização que os erros grandes levam. A equação é apresentada 3.6.

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2 \quad (3.6)$$

O RMSE também mede a magnitude média do erro, com uma pequena diferença para o MSE, que é a raiz quadrada. Assim a unidade de medida volta para a unidade real facilitando a interpretação da precisão do modelo. A fórmula é apresentada em 3.7.

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2} \quad (3.7)$$

3.4.3 R^2

Outra medida é o R^2 que mede o quão bem o modelo será capaz de prever os valores futuros. Na equação 3.8, o y são os valores reais, \hat{y} são os valores previstos, \bar{y} é a média dos valores reais e n é o número de amostras.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n-1} (y_i - \bar{y})^2} \quad (3.8)$$

3.5 Etapa 05: Visualização dos resultados

A etapa final da análise é a apresentação dos resultados. Nesse caso, será a previsão de vendas com uma antecedência mínima de cinco semanas, exportado para uma tabela no

Excel.

3.6 Materiais

Para a realização desta atividades foi utilizada a ferramenta Anaconda, que é uma distribuição da linguagem Python.

A biblioteca Pandas pode ser utilizada para a leitura e gravação de conjuntos de dados em diferentes formatos, tais como arquivos comma-separated values e de texto, planilhas em excel, entre outros. Ao ler os dados é criado um objeto DataFrame que facilita a manipulação de dados e possui indexação integrada. Com o DataFrame é possível realizar cálculos estatísticos básicos, operações nos conjuntos de dados, como mesclar e manipular valores.

A biblioteca Numpy foi utilizada para a manipulação das matrizes, operações de álgebra linear e transformação nos dados. Para a visualização de dados, foi utilizado a biblioteca Matplotlib que permite a elaboração de gráficos, histogramas e diagramas.

O IPython é um ambiente interativo que permite a visualização imediata dos resultados e realizar anotações. A biblioteca Scikit-learn possui um coleção de algoritmos e métodos implementados para análise de dados.

Flask é um microframework com código aberto que utiliza-se para o desenvolvimento de aplicações Web. A sua arquitetura baseia-se em adicionar os módulos conforme a necessidade do desenvolvimento.

Capítulo 4

Análise Exploratória e Preparação dos Dados

4.1 Análise Estatística do Dataset

Os dados fornecidos para o desenvolvimento deste trabalho são de uma fabricante internacional de componentes automóveis, localizada no nordeste de Portugal. O conjunto de dados refere-se às vendas realizadas no período de fevereiro de 2016 até janeiro de 2018.

O arquivo engloba 139.970 mil amostras com 14 colunas. A Tabela 4.1, apresenta uma amostra das cinco primeiras linhas do conjunto de dados antes do pré-processamento, entretanto alguns campos foram omitidos ou substituída as referências devido a sigilo e confidencialidade dos dados. Cada coluna apresenta as informações descritas, em que:

- Plant: Identificação da fábrica, sempre constante. Valor: 1361;
- Material: Referência interna do produto manufaturado. Possui 855 referências distintas;
- Purchase Order: Refere-se as ordens de compra. Sempre nulo, não apresentou nenhum dado;

- Material Description: Descrição do nome do produto por extenso. Possui 814 descrições distintas;
- Movement Type: Identificação do tipo de movimento do produto. Possui 2 valores distintos;
- Storage Location: Abreviação do local de armazenamento. Possui 6 valores distintos;
- Quantity: Quantidade do produto vendida ou devolvida;
- Order Unit: Unidade do pedido. Sempre nulo, não apresentou nenhum dado;
- Entry Date: Data em que foi realizada a venda. Possui 574 dias distintos;
- Time of Entry: Horário em que foi realizada a venda;
- Reference: Identificação única do produto no documento emitido pela fabricante. Possui 16014 valores distintos;
- Amount in LC: Valor total da quantidade na moeda local em que o sistema foi configurado;
- Material Document: Identificação dos produtos que foram vendidos ou devolvidos. Possui 16113 valores distintos;
- User name: Identificação do usuário. Possui 25 usuários distintos.

Tabela 4.1: Amostra dos dados originais fornecidos pela fabricante

	Plant	Material	Purchase Order	Material Description	Movement Type	Storage Location	Quantity	Order Unit	Entry Date	Time of Entry	Reference	Amount in LC	Material Document
0	AAQQ.0	007	NaN	Material 007	601.0	OU10	-90	NaN	2016-06-30	12:45:06	Z.0	-778.83	13
1	AAQQ.0	007	NaN	Material 007	601.0	OU10	-90	NaN	2016-06-30	12:40:44	J.0	-778.83	12
2	AAQQ.0	015	NaN	Material 015	601.0	OU10	-144	NaN	2016-06-30	19:20:12	H.0	-389.96	9
3	AAQQ.0	027	NaN	Material 027	601.0	OU10	-208	NaN	2016-06-30	19:20:12	H.0	-604.48	9
4	AAQQ.0	049	NaN	Material 049	601.0	OU10	-15	NaN	2016-06-30	21:45:05	Q.0	-48.61	14

Conforme analisado, optou-se por remover as colunas Purchase Order e Order Unit, por não terem nenhuma informação que poderia ser utilizado. Outra coluna removida foi a Plant, por apresentar valor constante.

Observou-se que a identificação da Reference era a mesma para as vendas e para as devoluções, desta maneira, foi possível identificar as ordens que foram vendidas e devolvidas. Isto é um ponto importante, devido a proposta de aumentar a precisão do modelo de previsão e diminuir a produção de produtos extras. A Tabela 4.2, apresenta que o MovementType de 601.0 é o valor de saída, assim a Quantity é negativa. Já o valor do MovementType 602.0 é de entrada, assim a Quantity é positiva. Foram identificadas 55 Reference iguais em que o Movement Type foi de saída e após um tempo de entrada, isso significa que estes produtos foram vendidos e logo em seguida devolvidos, deste modo, realizou-se a remoção destes dados do conjunto.

Tabela 4.2: Amostra dos dados de vendas e depois de devolução

Plant	Material	Purchase Order	Material Description	Movement Type	Storage Location	Quantity	Order Unit	Entry Date	Time of Entry	Reference	Amount in LC	Material Document	
0	AAQQ.0	001	NaN	Material 001	601.0	OU10	-90	NaN	2016-06-30	12:45:06	A.0	-778.83	1
1	AAQQ.0	002	NaN	Material 002	601.0	OU10	-90	NaN	2016-06-30	12:40:44	B.0	-778.83	2
2	AAQQ.0	003	NaN	Material 003	601.0	OU10	-144	NaN	2016-06-30	19:20:12	C.0	-389.96	3
3	AAQQ.0	004	NaN	Material 004	601.0	OU10	-208	NaN	2016-06-30	19:20:12	C.0	-604.48	3
4	AAQQ.0	005	NaN	Material 005	601.0	OU10	-15	NaN	2016-06-30	21:45:05	D.0	-48.61	4

Após isto, foi realizado o agrupamento pelas colunas Material, MaterialDescription e EntryDate, somando a Quantity, obtendo assim a vendas diárias de cada produto. Ao total esta nova tabela possuía 57.065 amostras, com quatro colunas. Foi realizado a operação matemática, valor absoluto, na coluna Quantity, tornando todos os valores positivos. A Tabela 4.3, apresenta 5 amostras das vendas diárias realizadas no decorrer do período de 01/02/2016 até 31/01/2018. Neste período foram produzidos 855 produtos diferentes pela empresa.

Tabela 4.3: Amostra dos dados de vendas diárias

Material	MaterialDescription	EntryDate	Quantity
001	Material 001	2016-02-01	200
001	Material 001	2016-02-08	200
001	Material 001	2016-02-22	200
001	Material 001	2016-03-07	200
001	Material 001	2016-03-14	200

A apresentação da correlação entre os dados Material e MaterialDescription, foram

aplicados na função para criar um Label Encoder, que transformou os dados não numéricos em dados numéricos, formando um único atributo, chamado Material_ID.

Realizou-se o agrupamento semanal, realizou-se uma análise cautelosa para que todas as amostras estivessem ao longo de uma semana completa. A atributo EntryDate foi utilizado para calcular o primeiro dia da semana, segunda-feira, chamado de Date.

Os dados foram novamente agrupados pelos atributos Material_ID e Date. Identificou-se que nem todos os produtos tinham o mesmo número de amostra por semana, pois nas semanas que não teve produção, não existia a informação. Para os Material_ID que não tinham vendas em determinada data, foram criadas amostras com o valor da Quantity zero, deixando o mesmo número iguais para todos os produtos. Assim o conjunto de dados ficou constituído com 853 produtos, em 104 semanas totalizando 88.712 mil amostras.

A Figura 4.1, apresenta o conjunto de dados agregados, nota-se uma sazonalidade nos meses de julho, agosto e dezembro, mais especificamente entre as semanas do ano 28 à 32 e 49 á 52. Esta queda das vendas, deve estar diretamente relacionadas ao período em que a fabricante tem por obrigação de fechar a fábrica, totalmente ou parcialmente, para as férias dos trabalhadores nos períodos de 15 de Julho a 31 de Agosto, por ser de até 21 dias consecutivos, e/ou durante as férias escolares de Natal, não excedendo 5 dias úteis consecutivos, conforme a cláusula 72, itens b e d do documento [28].

4.2 Agrupamento dos dados

A aplicação do método do agrupamento dos produtos descrito em 3.3.1, constatou-se que apenas duas classes estavam presentes no conjunto de dados sendo elas, a classe regular com 188 produtos e a classe irregular com 665 produtos.

A classe irregular representa 11% da quantidade de vendas totais de todos os produtos que a empresa realizou, enquanto o restante das vendas é representada pela classe regular.

Identificou-se que alguns produtos que foram classificados como irregulares, possuíam semanas regular de encomendas. Devido a este fato, criou-se uma tabela com 855 linhas e 104 semanas, em que o valor das célula 0 representava que o produto não teve encomendas

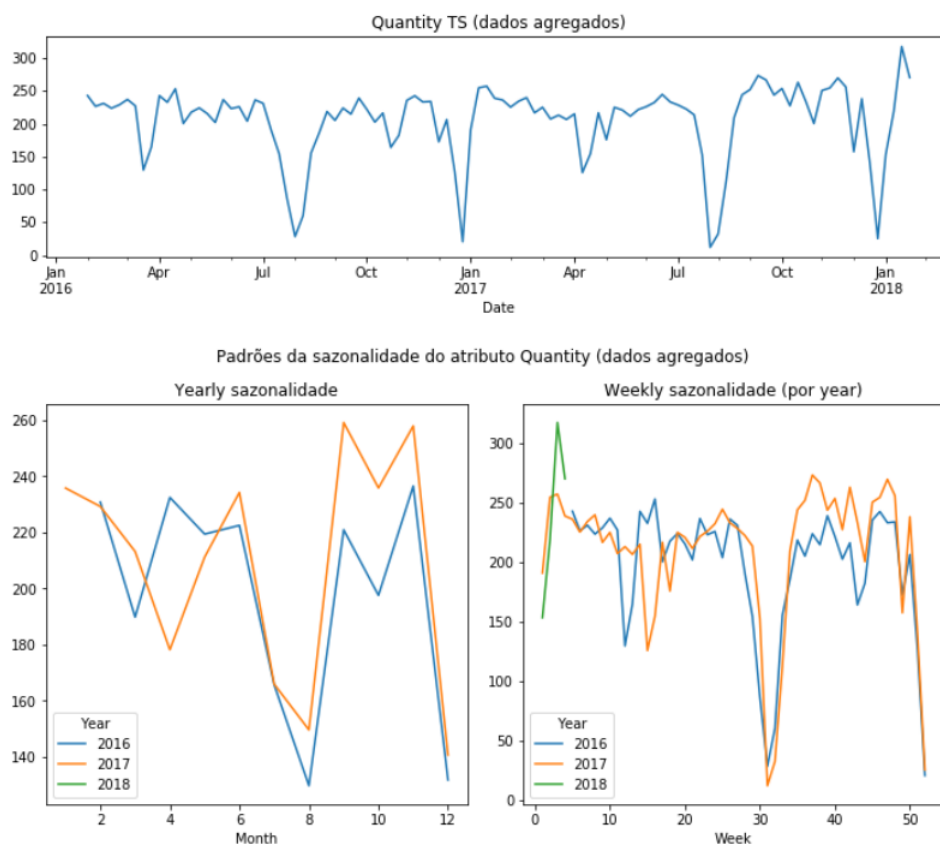


Figura 4.1: Sazonalidade dos dados agrupados da série temporal

naquela semana, enquanto o valor 1 representava que aconteceu encomendas.

Utilizou-se do algoritmo k-means para realizar o agrupamento dos produtos em 4 classes, seguindo o padrão descrito em 3.3.1.

A primeira classe, 0, contém 77 produtos, em que as encomendas variam entre 19 e 77 semanas. A percentagem de vendas totais é de 13%. A segunda classe, 1, contém 83 produtos, em que as encomendas variam entre 12 e 59 semanas. A percentagem de vendas é de 10%. A terceira classe, 2, contém 555 produtos, em que as encomendas variam entre 1 até 28 semanas. Esta classe contém os produtos que foram produzidos poucas vezes, normalmente um pedido especial de alguma empresa e os produtos descontinuados. A percentagem de vendas é de 2%. A quarta classe, 3, contém 138 produtos, em que as encomendas variam entre as semanas 54 e 102. São os produtos mais regulares, que têm vendas semanais constantes. Assim, é onde se encontra a maior representação de vendas que a empresa realizou, com uma percentagem de 75%.

A Tabela 4.4, apresenta a comparação entre as duas técnicas, descrita na seção 3.3.1 representada nas linhas e a aplicação da técnica k-means representada pelas colunas, em que a classe com identificação 3, agrupou todos os produtos regulares, 138 produtos, e a classe com identificação 2, contém todos os produtos irregulares ou descontinuados, com 555 produtos. Os outros dois identificadores de classe, 0 e 1, contêm a mistura entre as duas classes do método 3.3.1.

Tabela 4.4: Comparação entre as classificações dos produtos

	0	1	2	3	Venda (%)
Regular	41	9	0	138	11
Irregular	36	74	555	0	89
Venda (%)	13	10	2	75	100

Capítulo 5

Resultados e Discussões

A utilização de diferentes conjuntos de treinamento aconteceu pelos motivos de que os algoritmos precisaram de atributos diferentes para a realização do treinamento, por exemplo, o modelo Média Móvel Integrada Auto-regressiva Sazonal (SARIMAX) necessitava de uma série temporal com os dados ordenados. Para os modelo de RF foi necessário extrair algumas características referentes à data.

5.1 Conjunto de Treinamento - Tipo 01

Este conjunto de treinamento é composto pelos atributos `Material_ID`, `Year`, `Month`, `Week`, e o *target* que é a `Quantity`. Para isto o conjunto original foi dividido em 2 sub-conjuntos 10% para *test* e 90% para *train*, e realizado a normalização dos dados. Foi realizado o treinamento do modelo `RandomForestRegressor`, em que o resultados dos atributos mais importantes para a decisão é apresentado na Figura 5.1. Constata-se que o `Material_ID` é o atributo mais relevante para a decisão na floresta aleatória. Verificou que o modelo contém 10 árvores de decisão, que formaram a floresta aleatória, com um total de 8.872 folhas.

A Tabela 5.1, apresenta as métricas de avaliações do modelo, em que os resultados para os sub-conjuntos de treinamento e de teste, apresentaram uma R^2 de aproximadamente 97% e 85 % respectivamente. A Figura 5.2, apresenta uma comparação entre o `y_real`,

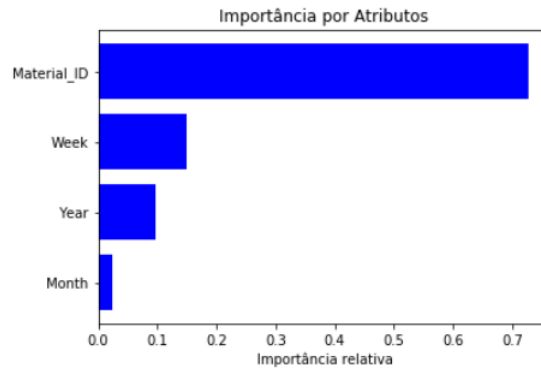


Figura 5.1: Importância dos atributos para o algoritmo floresta aleatória

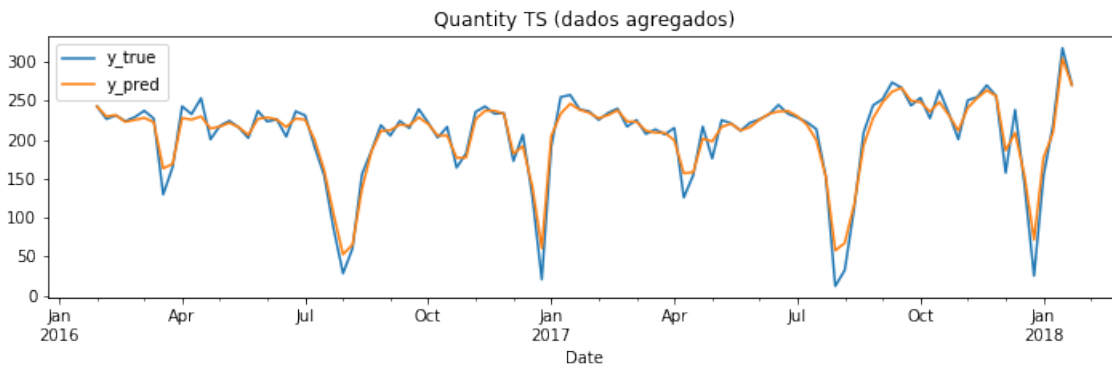


Figura 5.2: Comparação entre o y_{real} e o $y_{previsto}$ por meio do modelo RF

linha azul, e o $y_{previsto}$, linha laranja, os dados foram agrupados e realizado a média deles por semana de todos os produtos. Se nota que o modelo consegue identificar os padrões das séries temporais.

Tabela 5.1: Resultados das métricas de avaliação para o modelo de RF

	R-Squader	Mean Absolute Error	Root Mean Squared Error
Treinamento	0.974799	28.079867	130.18012
Teste	0.852687	69.956249	330.680791

5.2 Conjunto de Treinamento - Tipo 02

O modelo SARIMAX(p, d, q, s), como mencionados na seção 2.2.2, teve uma das seguintes configurações SARIMAX(1, 0, 0, 52) ou SARIMAX(0, 0, 0, 52). Com isto, realizou-se a

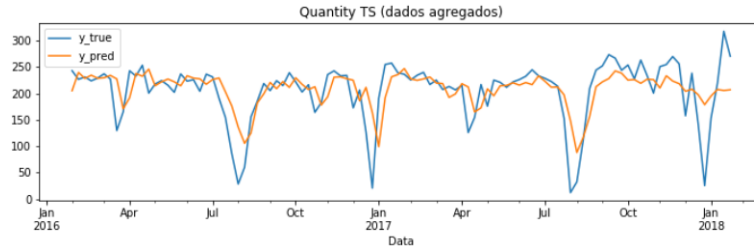


Figura 5.3: Comparação entre o y_{real} e o y_{previsto} obtido por meio do modelo SARIMAX

previsão para todos os produtos com a mesma configuração e com os resultados obtidos foram calculadas as métricas que são apresentadas na Tabela 5.2, nota-se que o conjunto *test* não teve uma boa resposta. A Figura 5.3, é notável que o modelo não conseguiu prever com tanta confiabilidade o comportamento das vendas dos produtos. Isto fica constatado pelo baixo valor do R^2 .

Tabela 5.2: Resultados das métricas de avaliação para o modelo de SARIMAX

	R-Squader	Mean Absolute Error	Root Mean Squared Error
Treinamento	0.858299	77.752127	308.478999
Teste	0.548395	133.945987	580.600020

5.3 Conjunto de Treinamento - Tipo 03

Com os resultados obtidos na Tabela 4.4, e a baixa importância dos produtos da coluna com o identificador 2, foi realizada a combinação entre os outros identificadores, em que a ordem foi a seguinte 0-A, 1-B, 3-C, 01-AB, 03-AC, 13-BC e 013-ABC, ou seja, realizou-se a combinação das colunas da Tabela 4.4 obtendo a lista dos produtos que pertenciam a cada classe. Os resultados obtidos são apresentados na Tabela 5.3, em que se observou uma melhoria nos resultados para o tipo C, com a combinação dos resultados em que, o R^2 aumentou e o MAE diminuiu, um bom sinal para o modelo que se espera encontrar para aplicar na empresa. Observando-se a métrica MAE no teste obteve-se um desvio aproximadamente de 75 a 270 unidades entre os produtos, em aproximadamente 10 semanas em que o modelo não conhecia os dados.

Tabela 5.3: Resultados das métricas de avaliação para o modelo de RF

	Treinamento			Teste		
	R-Squader	Mean Absolute Error	Root Mean Squared Error	R-Squader	Mean Absolute Error	Root Mean Squared Error
TipoA	0.978990	43.574911	133.585334	0.834241	128.023397	337.228069
TipoB	0.974476	34.224678	152.159194	0.689388	74.284212	250.057835
TipoC	0.977616	99.348963	231.701881	0.852998	270.826957	591.854025
TipoAB	0.971583	42.081782	155.212392	0.784749	119.655998	399.008583
TipoAC	0.976711	81.609975	210.744787	0.868531	218.903493	539.958559
TipoBC	0.977259	75.733943	208.442563	0.897246	204.788501	478.324194
TipoABC	0.975184	70.096943	201.247183	0.886182	188.441661	494.242138

5.4 Conjunto de Treinamento - Tipo 04

Para os algoritmos MLP e LSTM foi utilizado o conjunto com os atributos de datas: Year, Month e Week mais os atrasos das encomendas em cinco semanas e o Material_ID. A Tabela 5.4 apresenta uma amostra dos dados do conjunto, em que os parâmetros de data são da semana atual e t-1 são as venda da semana anterior, t-2 de duas semanas atrás, e assim por diante.

Tabela 5.4: Amostra do conjunto de treinamento do tipo 04

	Material_ID	Year	Month	Week	t-1	t-2	t-3	t-4	t-5
Date									
2016-03-07	5	2016	3	10	240.0	360.0	456.0	288.0	504.0
2016-03-14	5	2016	3	11	432.0	240.0	360.0	456.0	288.0
2016-03-21	5	2016	3	12	726.0	432.0	240.0	360.0	456.0
2016-03-28	5	2016	3	13	408.0	726.0	432.0	240.0	360.0
2016-04-04	5	2016	4	14	480.0	408.0	726.0	432.0	240.0

A Tabela 5.5 apresenta as métricas de avaliação do modelo MLP em que se observa uma pequena variação entre os resultados do conjunto de treinamento e de teste. Não houve melhoria nos erros de previsão conforme apresentado pelas métricas *MAE* e *RMSE* ao serem comparados com os valores das tabelas anteriores. O mesmo aconteceu com o algoritmo LSTM, em que não foram observadas melhorias, mas sim uma diminuição no coeficiente de determinação, R^2 . Em uma comparação entre os dois algoritmos, o MLP apresentou os melhores resultados com erros de previsão, MAE e RMSE, menores que os valores apresentados pelo modelo LSTM.

Tabela 5.5: Resultados das métricas de avaliação para o modelo de MLP e LSTM

	Treinamento			Teste		
	R-Squader	Mean Absolute Error	Root Mean Squared Error	R-Squader	Mean Absolute Error	Root Mean Squared Error
MLP	0.855600	214.173538	492.566256	0.851373	208.906001	490.093804
LSTM	0.577195	341.364286	849.326125	0.593557	331.294951	749.347785

Capítulo 6

Aplicação

Após identificar o melhor modelo, realizou-se o desenvolvimento de uma aplicação web para a utilização pelos usuários, podendo assim, realizar o carregamento dos arquivos sobre as vendas, em que a aplicação realizará o processamento do dados e exibirá os resultado por meio de gráficos ou saídas no formato excel.

Para o desenvolvimento da aplicação foi utilizada a linguagem de programação Python com o framework Flask e com a Ambiente de Desenvolvimento Integrado (IDE) Visual Studio Code¹.

6.1 Descrição dos Processos de Pré-Processamento

A Figura 6.1, apresenta as atividades realizadas durante o pré-processamentos dos dados. Após carregar os arquivos com a biblioteca Pandas, foi realizado a seleção das variáveis usadas neste trabalho de dissertação, seguido pelo processo para realizar o agrupamento dos produtos com o objetivo de identificar os produtos com demanda regular e por fim salvar em um arquivo todos os dados processados.

¹<https://code.visualstudio.com/>

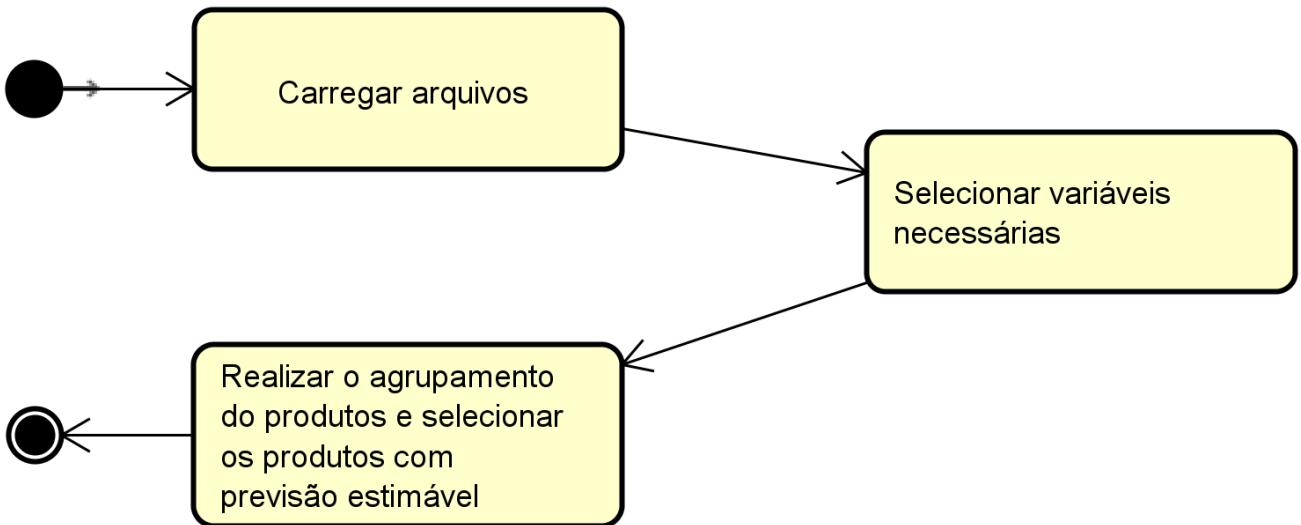


Figura 6.1: Diagrama com as atividades executadas no pré-processamento dos arquivos de dados brutos

6.1.1 Carregar arquivos

Para iniciar, é criado uma nova estrutura de dados tabulares bidimensional, onde foi inserido os dados brutos dos arquivos fornecidos pela fabricante. Nota-se que é removido qualquer linha que esteja com valor vazio na coluna Material.

```

1 df_vendas = pd.DataFrame()
2 for vendas_file in os.listdir(path_raw_vendas):
3     df_vendas = df_vendas.append(pd.read_excel(path_raw_vendas + vendas_file))
4     df_vendas = df_vendas.dropna(subset=['Material'], how='all')
  
```

Algoritmo 6.1: Leitura dos dados brutos

6.1.2 Selecionar variáveis necessárias

Conforme explicado na Seção 4.1, nem todos os atributos foram utilizado para realizar a previsão, assim o seleção destes atributos é apresentando no Algoritmo 6.5. A função groupby, agrega todas as informações semelhantes realizando a soma. Com isso, é obtido a Quantidade total vendida no dia, mais esta quantidade está negativa, assim é realizado

o *abs*, valor absoluto do atributo.

```
1 df_vendas = df_vendas.groupby(['Material', 'MaterialDescription',  
2                               'EntryDate'])['Quantity'].sum().reset_index()  
3 df_vendas['Quantity'] = abs(df_vendas['Quantity'])
```

Algoritmo 6.2: Somatório da quantidade agrupado pelos atributos Material, Material Description e Entry Date

6.1.3 Realizar o agrupamento dos produtos e selecionar os produtos com previsão estimável

Para realizar o agrupamento dos dados foi criando um dataframe que contém as informações referentes aos produtos que foram vendido ou não naquela semana. Com esse dataframe foi realizado o treinamento da técnica k-means, obtendo assim as classes semelhantes. O Algoritmo 6.3 apresenta os passos realizados.

```
1 from sklearn.cluster import KMeans  
2  
3 df_cluster = pd.DataFrame(index=df_vendas.Material_ID.unique(), columns=df_vendas.Date.unique())  
4 for index in df_vendas.Material_ID.unique():  
5     for col in df_vendas.Date.unique():  
6         df_cluster.at[index, col] = np.where(df_vendas[(df_vendas['Material_ID'] == index)  
7  
8 kmeans = KMeans(n_clusters = 4, init = 'random', max_iter=400, random_state=45)  
9 kmeans.fit(df_cluster)  
10 labels = kmeans.labels_  
11  
12 df_kmeans = pd.DataFrame()  
13 df_kmeans['Material_ID'] = df_cluster.index  
14 df_kmeans['Label'] = labels  
15
```

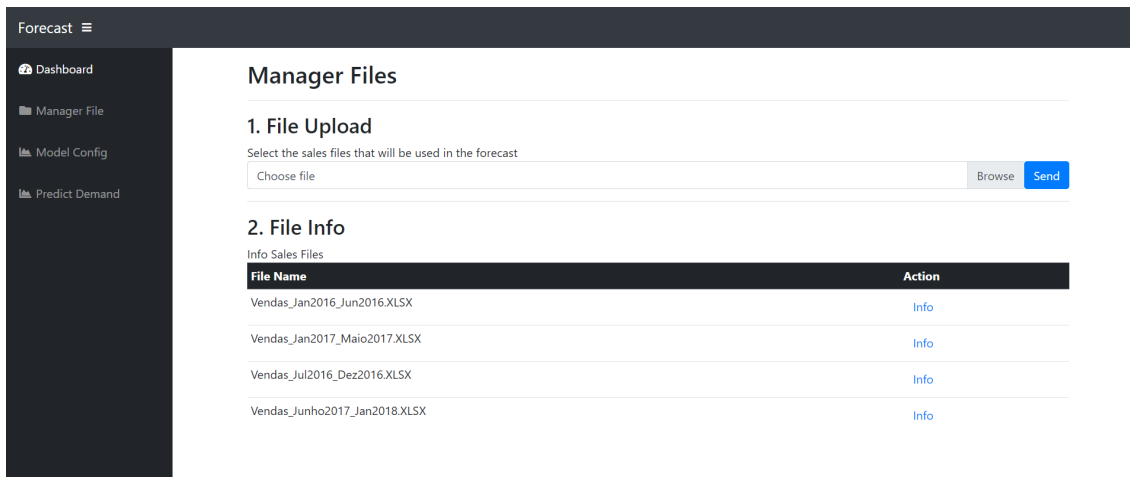


Figura 6.2: Tela do gerenciamento dos arquivos brutos do conjunto de dados

```
16 list_material_id = df_kmeans[df_kmeans['Label'].isin([0,1,3])].index
```

Algoritmo 6.3: Etapa do agrupamento dos produtos e seleção dos produtos com demanda regular

A Figura 6.2 demonstra os passos descritos anteriormente, com o processo para selecionar os arquivos e o gerenciamento dos mesmos.

6.2 Descrições sobre os Processos do Treinamento do Modelo

Após os ajustes no conjunto de dados foi necessário projetar como o modelo é treinado. Assim são apresentadas as etapas por meio do diagrama de atividades na Figura 6.3. Basicamente o processo é composto de 3 principais fases: a separação da variável dependente, Quantity e das variáveis independentes.

Após essa separação é realizado a divisões. Por fim, é realizado o treinamento do modelo e salvo o mesmo para uma utilização futura.

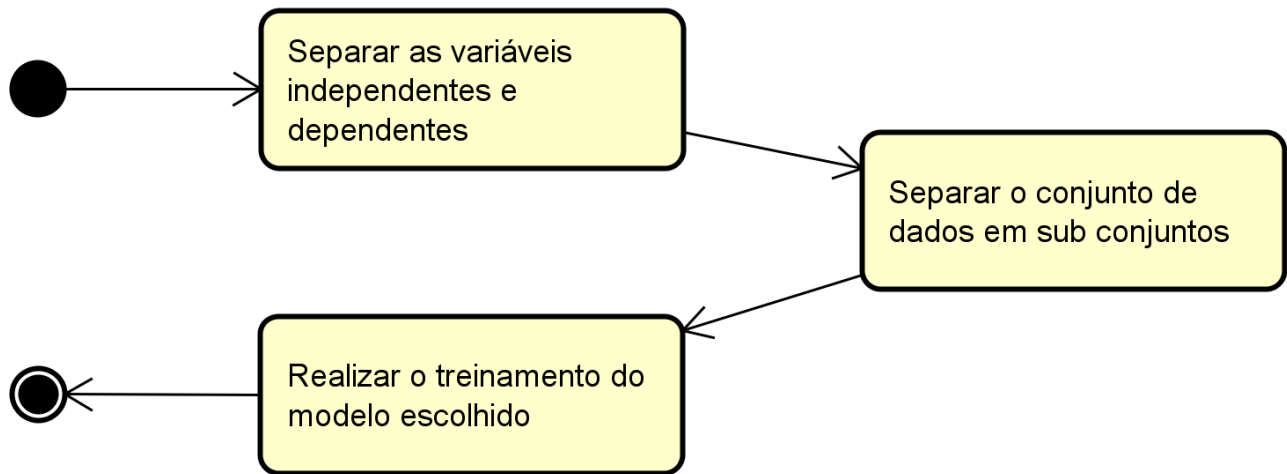


Figura 6.3: Diagrama com as atividades executadas no treinamento do modelo

6.2.1 Separação das variáveis independentes e dependentes e modelação do conjunto de dados nos sub-conjuntos de treinamento e teste

O Algoritmo 6.4, apresenta o separação do conjunto de dados em duas variáveis X e y. A variável X contém as dados independentes, como o mês, ano e semana do ano,; enquanto a variável y contém os dados alvo que é a Quantity.

O segundo passo, foi realizar uma transformação que tornou o conjunto com média zero e dentro de uma escala de variação entre os valores 0 e 1.

De seguida foi separado o conjunto em 10% para teste e 90% para o treinamento.

```

1 X = df_vendas.drop(columns=['Quantity']).set_index('Date')
2 y = df_vendas.set_index('Date')[["Quantity"]].copy()
3 def(X, y, porc=.10):
4     scx=MinMaxScaler()
5     scy=MinMaxScaler()
6     scx.fit(X)
7     scy.fit(y)
8     Xm=pd.DataFrame(scx.transform(X), columns=X.columns)
9     Xm.index = X.index
  
```

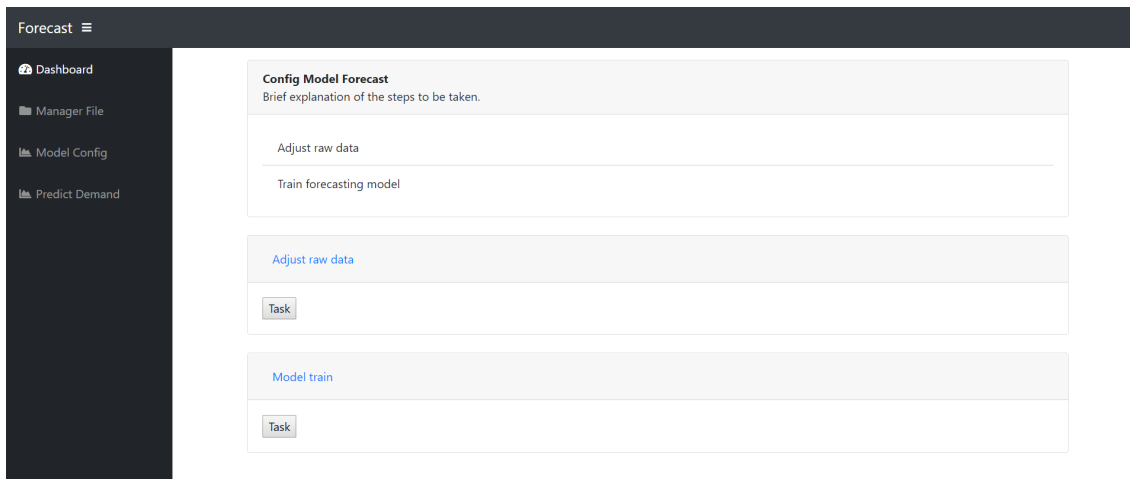


Figura 6.4: Tela das tarefas de previsão de treinamento do modelo

```

10  ym=pd.DataFrame(scy.transform(y), columns=y.columns)
11  ym.index = y.index
12  # Separa os conjuntos em teste e validacao com 10%
13  Xt, Xv, yt, yv= train_test_split(Xm,ym, test_size=porc, random_state =302)

```

Algoritmo 6.4: Procedimento de separação das variáveis e a criação dos conjuntos

6.2.2 Realizar o treinamento do modelo escolhido

Após terem sido criados os conjuntos de treinamento e teste, o conjunto de treinamento foi usado para realizar o treinamento do modelo e em seguida salvo no disco.

```

1  model_RF= RandomForestRegressor()
2  model_RF.fit(Xt, yt)
3  # save the model to disk
4  filename = 'finalized_model.sav'
5  joblib.dump(model, path_model + filename)

```

Algoritmo 6.5: Treinamento do modelo e salvamento no disco local

A Figura 6.4 apresenta a página de interação com os algoritmos descritos nesta seção.

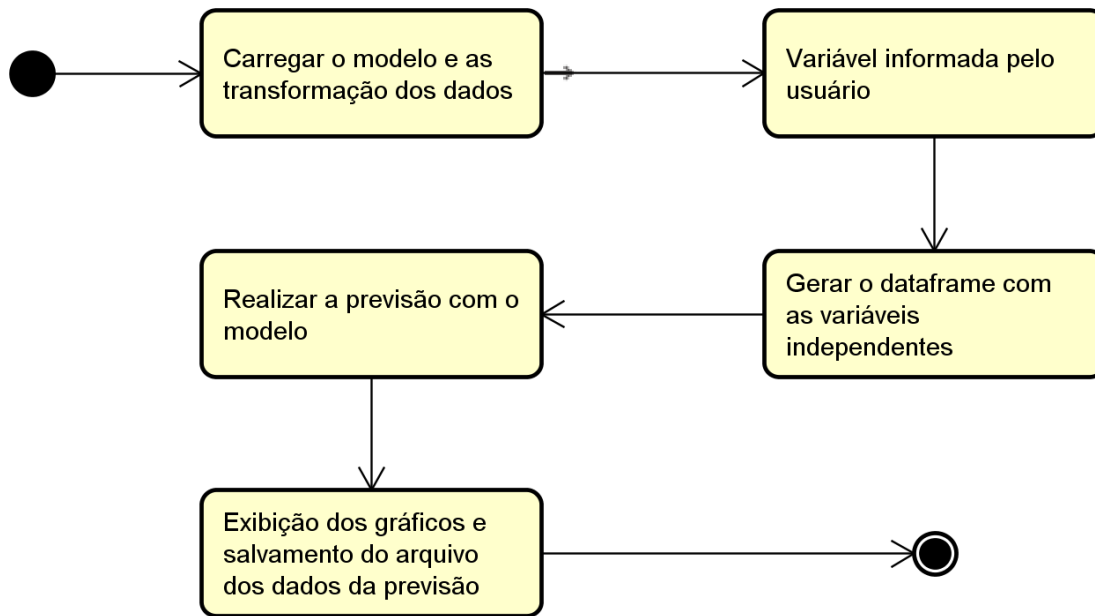


Figura 6.5: Diagrama com as atividades executadas na previsão com o modelo

6.3 Descrições sobre os Processos de Previsão do Modelo

Após a conclusão das tarefas das seções 6.1 e 6.2, foi possível realizar a previsão para valores futuros da demanda. A Figura 6.5 apresenta os processos que foram necessários para realizar a previsão e obter os resultados especulativos.

6.3.1 Carregar o modelo e as transformação dos dados

Este processo acontece pelo fato de não ser viável realizar um treinamento do modelo toda a vez que for solicitado uma previsão. Assim, neste processo é carregado as informações essenciais que são o modelo construído e as variáveis de transformações dos dados que foram utilizadas.

6.3.2 Variável informada pelo usuário

Para a realização da previsão os parâmetros que foram solicitados ao usuário foram:

- Ano:
- Semana do ano:
- Numero de semanas a serem previstos:

6.3.3 Realizar a previsão com o modelo

Após o treinamento do modelo é esperado obter resultados sobre o rumo dos produtos manufaturados pela empresa e assim realizar as tomadas de decisão. Para tal objetivo, foi realizado a criação de uma função em Python.

```

1 def fun_forecast(year, weekofyear, num_forecast):
2     import datetime
3     dt= datetime.datetime(year,1,1)
4     # Check if it's the first day of the week
5     if dt.weekday() != 0:
6         dt = dt - timedelta(days=dt.weekday())
7     # Checks if it is the first week of the year
8     if dt.isocalendar()[1] != 1:
9         dt = dt + timedelta(days=7)
10    # Calculates the date of the forecast week requested
11    dt = dt + timedelta(days=7*weekofyear)
12    df_aux_pdp = pd.DataFrame(data=list_material_id)
13
14    # For each week the forecast of the products is
15    # carried out and stored in the other dataframe
16    for x in range(num_forecast):
17        dt_previsao = dt + timedelta(days=7*x)
18        aux_df = pd.DataFrame( data=list_material_id)
19        aux_df[ 'Year' ] = dt_previsao.year
20        aux_df[ 'Month' ] = dt_previsao.month
21        aux_df[ 'Week' ] = dt_previsao.isocalendar()[1]
22
23        Xm=pd.DataFrame(scx.transform(aux_df), columns=aux_df.columns)
24        predicted_train = model_RF.predict(Xm)

```

```

25
26     aux_df[ 'prev' ] = scy.inverse_transform(predicted_train.reshape(-1, 1))
27     aux_df[ 'prev' ] = aux_df[ 'prev' ].apply(lambda x : math.floor(x))
28
29     df_aux_pdp[ '' .join([ 'W+', str(x+1)])] = aux_df[ 'Week' ]
30     df_aux_pdp[ '' .join([ 'Qtd_W+', str(x+1)])] = aux_df[ 'prev' ]

```

Algoritmo 6.6: Processo de criação das variáveis independentes e processo de previsão

6.3.4 Exibição dos gráficos e salvamento do arquivo dos dados da previsão

No apêndice, Tabela C.1 encontra-se as estimativas realizadas pelo modelo para o ano de 2018, entre as semanas do ano 5 e 9. Os produtos/materiais previstos são os de demanda regular totalizando 298 produtos. Nota-se que os dados conhecidos eram até à semana 4 do ano de 2018, sendo assim estes resultados não foram comparados com a real demanda que a fabricante teve durante esse período.

Por fim a exibição na aplicação web foi realizado por meio do Dash ², assim foi possível a integração entre os framework Flask e Dash, criando interfaces interativas para a visualização dos produtos previsto por meio de gráficos. A Figura 6.6 apresenta a dashboard com os gráficos de barras das semanas para as quais foi realizada as previsões e ainda foi disponibilizado filtro para todos os elementos, facilitando a visualização nos gráficos.

²<https://plot.ly/products/dash/>

Dashboard Previsão

FILTER ROWS									
<input type="checkbox"/>	Material	MaterialDescription	Qtd_W+1	Qtd_W+2	Qtd_W+3	Qtd_W+4	Qtd_W+5	W+1	W+
<input type="checkbox"/>	001	Material 001	1454	1559	1559	1559	1559	5	6
<input type="checkbox"/>	002	Material 002	2777	2928	3091	3091	2934	5	6
<input type="checkbox"/>	003	Material 003	2755	2833	2822	2956	2912	5	6
<input type="checkbox"/>	004	Material 004	1205	1205	1205	1205	1205	5	6
<input type="checkbox"/>	005	Material 005	1888	1888	1888	1888	1888	5	6
<input type="checkbox"/>	006	Material 006	1349	1349	1349	1349	1349	5	6
<input type="checkbox"/>	007	Material 007	1651	1651	1651	1651	1651	5	6
<input type="checkbox"/>	008	Material 008	469	530	449	449	449	5	6
<input type="checkbox"/>	009	Material 009	1980	1980	1980	1980	1980	5	6
<input type="checkbox"/>	010	Material 010	235	182	110	220	278	5	6

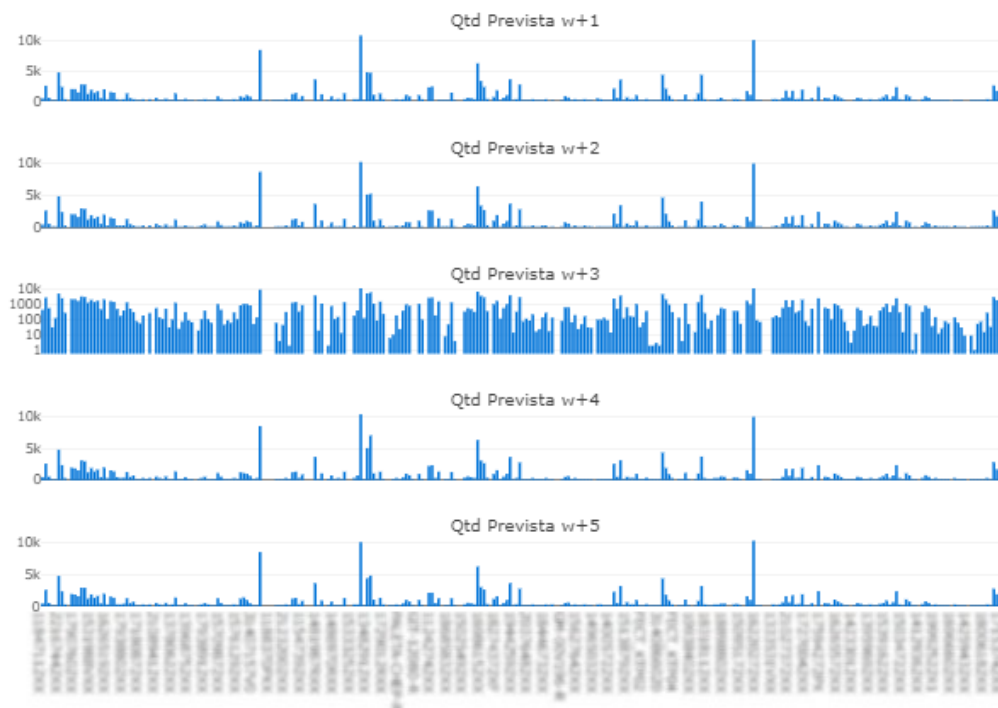


Figura 6.6: Dashboard dos produtos previstos com os gráficos e filtros

Capítulo 7

Conclusões e Trabalhos Futuros

7.1 Conclusão

A previsão ainda continua sendo um desafio a ser estudado. As limitações referentes à conjuntos pequenos de dados são um dos problemas que requer a pesquisa de novas soluções. A aplicação de técnicas de *Machine Learning* (ML) podem contribuir para uma melhoria na previsão dos dados.

A identificação dos produtos que são produzidos em larga escala é um ponto importante. Verificando o resultado entre os dados dos conjuntos que possuíam todos os dados e o conjuntos já com os dados filtrados, fica notável a importância do agrupamento. Devido ao fato de não ter dados rotuladas foi procurado métodos na literatura e por meio das técnicas de agrupamento do aprendizado de máquina, assim foi selecionado a melhor forma em que foi possível realizar a remoção de produtos que foram produzidos poucas semanas ou produtos descontinuados ao longo do ano, tudo isso automaticamente, sem a interferência humana.

Em uma avaliação geral do conjunto de dados testado, o melhor algoritmo foi o RF, com ele foi possível atingir um precisão de aproximadamente 88% da variável prevista, Tabela 5.3 - Tipo ABC.

Os resultados do algoritmo MLP são promissores, Tabela 5.5, apresentando uma precisão de aproximadamente 85%, demonstrando um bom R^2 .

O algoritmo de SARIMAX, não apresentou bons resultados para o conjunto de teste, isso significa que o mesmo teve uma baixa correlação entre os valores previstos e os valores reais, com um R^2 de aproximadamente 55%.

Depois de testar algumas soluções e realizar a análise em busca do melhor conjunto de dados e do melhor algoritmo para resolver este problema, foi selecionado o conjunto de dados do tipo 03 e o algoritmo de aprendizado de máquina *Random Forest* (RF). Por fim, o desenvolvimento de uma aplicação para facilitar a utilização pelos usuários. As vantagens são o ganho de informações que é possível com a utilização dessa aplicação, proporcionando mais tempo para realizar uma análise humana nos dados gerados, realizando ajustes quando for necessário.

7.2 Trabalhos Futuros

Como possíveis trabalhos futuros, pode-se apontar:

- Reunir as informações históricas de no mínimo 5 anos para um novo conjunto de dados e assim realizar a extração de características e modelos de previsão.
- Busca de incorporação de dados que possuem Influências nas vendas, como condições do mercado de atuação, condições climáticas, e assim por diante.
- Aprendizagem por reforço, em que desenvolvimento de um validação automática e correção nos parâmetros com o passar do tempo.
- Desenvolvimento com a utilização de dados em tempo real ou por semana, interligando as aplicações de previsão juntamente com o Sistema Integrado de Gestão Empresarial (ERP) utilizado pela empresa.
- Combinação das técnicas de previsão de forma a criar uma rede com vários previsores, combinando os melhores modelos em um só, basicamente como a RF faz, em

que é um conjunto de árvores de decisão onde é realizado uma votação para escolher o resultado.

- Implementação de previsão em varias etapas, pois a previsão atual foi desenvolvida em um passo sendo necessário repetir N vezes para obter a quantidade desejada.
- Com mais amostras no conjunto de dados, explorar as componentes de tendência e sazonalidade na procura para encontrar os melhores parâmetros para as técnicas tradicionais baseadas em modelos matemáticos.

Bibliografia

- [1] J. Peinado e A. R. Graeml, “Administração da produção”, *Operações industriais e de serviços. Unicenp*, 2007.
- [2] E. T. Jaynes, “The minimum entropy production principle”, *Annual Review of Physical Chemistry*, vol. 31, n.º 1, pp. 579–601, 1980.
- [3] J. T. Mentzer e M. A. Moon, *Sales forecasting management: a demand management approach*. Sage, 2004.
- [4] M. A. Moon et al., “Breaking down barriers to forecast process improvement”, *Foresight: The International Journal of Applied Forecasting*, vol. 4, pp. 26–30, 2006.
- [5] S. Makridakis, E. Spiliotis e V. Assimakopoulos, “Statistical and Machine Learning forecasting methods: Concerns and ways forward”, *PloS one*, vol. 13, n.º 3, e0194889, 2018.
- [6] D. Simchi-Levi e E. Simchi-Levi, “Inventory optimization: the last frontier”, *IT matters*, vol. 3, 2004.
- [7] S. Chopra e P. Meindl, “Gerenciamento da Cadeia de Suprimentos: Estratégia, Planejamento, e Operação”, em *Gerenciamento da Cadeia de Suprimentos: Estratégia, Planejamento, e Operação*, 2003.
- [8] G. I. KENDALL, “Visão Viável-Transformando o Faturamento em Lucro Líquido”, *Porto Alegre: Bookmann*, 2007.
- [9] D. Hutchins, *Just in time*. Gower Publishing, Ltd., 1999.

- [10] H. L. Corrêa e C. A. Corrêa, *Administração de Produção E Operações: Manufatura E Serviços: Uma Abordagem Estratégica*. Editora Atlas SA, 2000.
- [11] L. M. Sanches e O. F. Lima Jr, “Hockey stick phenomenon: supply chain management challenge in Brazil”, *BAR-Brazilian Administration Review*, vol. 11, n.º 3, pp. 264–283, 2014.
- [12] H. Takeda, *The Synchronized Production System: Going Beyond Just-In-Time Through Kaizen*. Kogan Page, Limited, 2006, ISBN: 0749447656.
- [13] E. A. Coimbra, *Total Flow Management: Achieving Excellence with Kaizen and Lean Supply Chains*. Kaizen Institute, 2009.
- [14] R. J. Hyndman, A. V. Kostenko et al., “Minimum sample size requirements for seasonal forecasting models”, *Foresight*, vol. 6, n.º Spring, pp. 12–15, 2007.
- [15] P. A. Morettin e C. Toloi, “Análise de séries temporais”, em *Análise de séries temporais*, 2006.
- [16] A. L. Bruni e R. Famá, “Eficiência, previsibilidade dos preços e anomalias em mercados de capitais: teoria e evidências”, *Caderno de Pesquisas em Administração*, vol. 1, n.º 7, pp. 71–85, 1998.
- [17] D. F. Silva, “Classificação de séries temporais por similaridade e extração de atributos com aplicação na identificação automática de insetos”, tese de doutoramento, Universidade de São Paulo.
- [18] J. O. Giacomantone, “Ressonância magnética funcional com filtragem pela difusão anisotrópica robusta.”, tese de doutoramento, Universidade de São Paulo, 2005.
- [19] R. J. Hyndman e G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.
- [20] V. Zocca, G. Spacagna, D. Slater e P. Roelants, *Python Deep Learning*, 2017.
- [21] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*. Chapman e Hall/CRC, 2012.

- [22] S. Russel e P. Norvig, “Inteligência artificial”, *Editora Campus*, p. 26, 2004.
- [23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot e E. Duchesnay, “Scikit-learn: Machine Learning in Python”, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [24] S. Hochreiter e J. Schmidhuber, “Long short-term memory”, *Neural computation*, vol. 9, n.º 8, pp. 1735–1780, 1997.
- [25] J. Zhang, Y. Zhu, X. Zhang, M. Ye e J. Yang, “Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas”, *Journal of Hydrology*, vol. 561, pp. 918–929, 2018.
- [26] H. Cuesta e S. Kumar, *Practical data analysis*. Packt Publishing Ltd, 2016.
- [27] A. A. Syntetos, J. E. Boylan e J. Croston, “On the categorization of demand patterns”, *Journal of the Operational Research Society*, vol. 56, n.º 5, pp. 495–503, 2005.
- [28] AIMMAP, SINDEL, FETESE e SITESC, *Contrato Colectivo de Trabalho para a Indústria da Metalurgia e Metalomecânica*, http://www.aimmap.pt/fotos/editor2/cct_2010.pdf, 2010.

Apêndice A

Escola de Verão & Simpósio de Dupla Diplomação 2018



Predição de Séries Temporais em Produção Industrial

(✉) D. Folletto¹, P. Matos², A. Gavioli³

¹ Mestrado em Sistemas de Informação, Escola Superior de Tecnologia e Gestão de Bragança, Instituto Politécnico de Bragança, Portugal.

¹ Bacharelado em Ciência da Computação, UTFPR - Câmpus Medianeira, Universidade Tecnológica Federal do Paraná, Brasil.

douglasmenegol1@hotmail.com

² Departamento de Informática e Comunicações, Escola Superior de Tecnologia e Gestão de Bragança, Instituto Politécnico de Bragança, Portugal.

pmatos@ipb.pt

³ Departamento de Ciência da Computação, UTFPR - Câmpus Medianeira, Universidade Tecnológica Federal do Paraná, Brasil.

alan@utfpr.edu.br

Resumo

A utilização dos processos de previsões de resultados ajudam empresas a realizar seu planejamento, definir metas, detectar anomalias, estimar a procura ou prever as necessidades produtivas de logística operacional, nomeadamente das encomendas de matéria prima, entre outros. Assim, a utilização de métodos preditivos tornam-se cada vez mais uma ferramenta fundamental para auxiliar nos processos de decisões empresariais, principalmente quando por meio destes, é possível obter previsões confiáveis e de elevada qualidade.

As previsões são obtidas por meio de dados históricos da empresa, muitas vezes sob a forma de séries temporais, que descrevem os dados ordenados no tempo.

Nesse contexto, essa dissertação de mestrado tem como objetivo principal, a criação de um modelo de previsão que tendo por base o histórico de encomendas dos clientes, que permita auxiliar no processo de definição do plano de produção, com uma antecedência mínima de cinco semanas.

A ferramenta utilizada para esse propósito é o *Prophet*, que utiliza modelos tradicionais de previsão, como Holt Winters e Box & Jenkins, para a partir de séries temporais, criar modelos de previsões. Realizou-se com esta o treinamento de um modelo de previsão com os dados reais, tendo-se alcançado um coeficiente de determinação de 78% que indica que o comportamento da variável dependente é altamente explicado pelo comportamento da variável independente.

Na Figura 1, a linha azul representa os dados reais e a linha vermelha os dados previstos. Para este caso, a previsão foi realizada com cinco semanas de antecedência, conforme definido nos requisitos. No eixo y estão as quantidades e no eixo x as datas em meses. A raiz do erro quadrático médio, RMSE do inglês, foi de 22.064,81 quantidades. Essa medida indica a diferença entre os valores previstos e os valores reais.

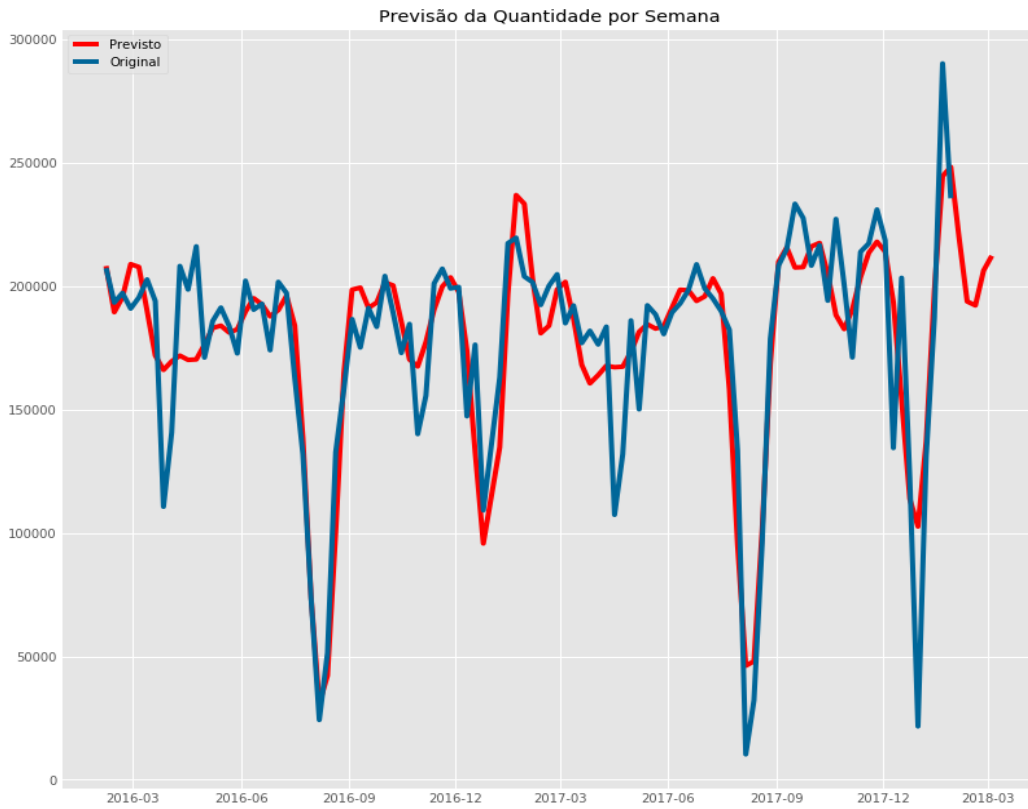


Figura 1: Previsão da quantidade por semanas do ano.

Não obstante o modelo ainda estar em validação, o objetivo no decorrer desse trabalho de dissertação é transpô-lo para ser uma solução autônoma e de utilização real para a empresa.

Apêndice B

9º MEDITEC – Medianeira in Technology

Predição de Série Temporal Aplicada no Processo Produtivo

Douglas Menegol Folletto¹, Paulo Jorge Teixeira Matos², Alan Gavioli³

¹ Instituto Politécnico de Bragança
Bragança, Portugal

² Departamento de Informática e Comunicações – Instituto Politécnico de Bragança
Bragança, Portugal

³ Departamento de Ciência da Computação – Universidade Tecnológica Federal do Paraná
Medianeira, Brasil

douglasmenegoll@hotmail.com, pmatos@ipb.pt, alan@utfpr.edu.br

Abstract. *The use of predictive methods is becoming a fundamental tool to aid in business decisions. So with the Facebook Prophet tool, you can create forecast models for time series that estimate future orders.*

Resumo. *A utilização de métodos preditivos estão se tornando uma ferramenta fundamental para auxiliar nas decisões empresariais. Assim, com a ferramenta Facebook Prophet, é possível criar modelos de previsão para séries temporais que estimam as encomendas futuras.*

1. Introdução

Em 1980, Jaynes foca em um objetivo fundamental, com a informação disponível, quais serão as melhores previsões obtidas sobre os fenômenos observáveis? [Jaynes 1980]. Surge o conceito da análise preditiva que é utilizada para a identificação dos padrões nos dados e então indicar tendências, mudanças, hábitos de consumo, entre outras coisas mais.

Normalmente o processo produtivo não fornece uma rápida resposta a demanda, isso significa que as empresas não podem aguardar que seus clientes manifestem interesse pelo produto para então iniciar sua produção, exceto em empresas que trabalham com o processo produtivo sob demanda, conforme [Slack et al. 2009]. Com isso, o processo produtivo inicia-se antes de conhecer a real necessidade do mercado, tornando-se necessário a utilização de ferramentas que auxiliem nesse processo, por exemplo as previsões.

A previsão de vendas é uma estimativa de quanto será vendido por uma empresa num determinado período de tempo. As contribuições da previsão de vendas no processo do planejamento corporativo é amplamente necessário e reconhecido [Mentzer and Moon 2004]. Isso é muito importante na logística operacional do processo produtivo, pois fornece uma orientação, evitando assim atrasos de entrega, perda de vendas, elevado custo da produção ou os custos com estoque.

Uma previsão de demanda de alta qualidade, permite que a empresa forneça aos seus clientes os produtos e serviços, quando e onde eles desejarem [Moon et al. 2006]. Com isso, a procura para melhorar os processos de previsão de demanda tornou-se necessário, devido ao auxílio fornecido no planejamento da cadeia de suprimentos, fornecimento futuro, estratégias de vendas e assim por diante.

2. Materiais e Métodos

Série temporal pode ser definida com um conjunto de amostra ordenadas cronologicamente. As observações do intervalo de tempo de séries temporais podem variar por dia, semana, mês, ano, entre outros. Alguns trabalhos que foram desenvolvidos utilizando séries temporais utilizam dados como, ações da bolsa de valores, imagem de análise, entre outros [Hyndman and Athanasopoulos 2018].

A estacionariedade expressa que a série temporal está se desenvolvendo no tempo aleatório em torno da média constante, tornando estável. Na maioria das vezes, a séries apresentam alguma forma de não-estacionariedade. Uma das formas de verificar a estacionariedade é por meio de gráficos ou de testes estatísticos, como Dickey-Fuller [Costa and Baidya 2001].

Os três principais padrões das séries temporais são a tendência, sazonalidade e ciclo. A tendência indica o comportamento da série ao longo do tempo, em que pode ser constante, linear ou quadrática. A sazonalidade em uma série são as oscilações que sempre acontecem em um determinado período do dia, do mês, da semana ou do ano. Um ciclo indica as alterações que não possuem períodos regulares, e que podem ser representadas por ondas longas, em torno da tendência [Hyndman and Athanasopoulos 2018].

Dentre os principais interesses no estudo de séries temporais estão a análise e modelagem da série temporal e a previsão na série temporal. O primeiro busca-se descrever a série, encontrar as características mais importantes e possível relações com outras séries. Na previsão espera-se com os valores históricos da série, estimar previsões de curto prazo, *forecast*, do inglês. O número de instancias futuras que serão previstas é conhecido como horizonte de previsão.

2.1. Dados e pré-processamento

O registro de dados de vendas diárias correspondentes ao período de fevereiro de 2016 até janeiro de 2018, foram obtidos em contato com uma fabricante internacional de peças automotivas, na região nordeste de Portugal.

Com um total de 139.970 registro de vendas e uma produção de 860 materiais durante o período fornecido. Os dados contém as informações sobre a fábrica, o material e sua descrição, o tipo de movimento, o local de armazenagem, a quantidade, o data e hora, a referência, o valor total na moeda local e o documento do material.

A Figura 1, apresenta as etapas que foram realizadas durante o pré-processamento dos dados brutos produzindo o arquivo de vendas semanas. Na etapa 1 foi realizado a leitura dos 4 arquivos em excel fornecidos pela fábrica e removido as linhas vazias que não possuíam as informações do material, pois eram as linhas de somatório de todas as quantidade vendidas e valor que estavam no arquivo do excel.

Na etapa 2, foi realizado a remoção das devoluções de produtos que estavam aparecendo nos dados fornecidos, isso foi possível pelo somatório da quantidade agrupamento pela referência, onde foi identificado que os valores se anulavam, ou seja, o somatório da referência apresentavam o valor zero. Depois foi realizar o agrupamento dos documento do material que possuíam valores com a quantidade positiva, que segundo a empresa não eram produtos que foram vendido, e assim realizado a remoção do dados.

Por fim na etapa 3 foi realizado o somatório da quantidade agrupados por material, descrição e data, seguido pelo cálculo do primeiro e do último dia da semana para cada amostra. Após isso foi realizado a remoção das amostras que não completaram uma semana inteira. Isso foi necessário para não influenciar durante a construção do modelo e o último passo foi realizado o somatório da quantidade agrupamento pelo primeiro dia da semana, formando a série temporal que foi analisada. Após todo o pré-processamento, a série temporal ficou com 104 amostras.

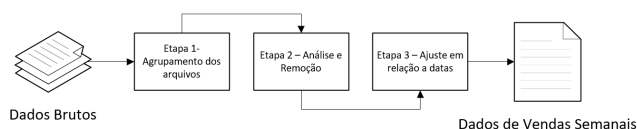


Figura 1. Etapas realizadas no pré-processamento

2.2. Facebook Prophet

O Phophet é uma biblioteca desenvolvida para auxiliar na criação de previsões de séries temporais, [Taylor and Letham 2018]. O modelo utilizado é apresentado na Equação 1, no qual as componentes são:

- $g(t)$ é o crescimento linear ou logístico das tendências.
- $s(t)$ são as mudanças periódicas, ou seja, sazonalidade.
- $h(t)$ é o efeito dos feriados, fornecidos durante o treinamento.
- $e(t)$ são os resíduos, as alterações não explicadas pelo modelo.

$$y(t) = g(t) + s(t) + h(t) + e(t) \quad (1)$$

Utilizando o tempo como variável independente, é realizado combinar de várias funções lineares e não lineares no tempo como os componentes. A sazonalidade da modelagem no Prophet é aditiva ou multiplicativa, a mesma abordagem utilizada para a suavização no modelo de Holt-Winters. Essa biblioteca realiza ajuste na curva do problema de previsão, sem olhar explicitamente para a dependência baseada no tempo de cada observação da série temporal.

2.3. Medidas de Avaliação

Para determinar o valor do coeficiente de determinação, R^2 , foi utilizado a Equação 2, em que a melhor pontuação é 1. O numerador da fração corresponde ao somatório dos quadrados do erro residual, enquanto o denominador corresponde ao somatório total dos quadrados, conforme [Damodar et al. 2004].

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n-1} (y_i - \bar{y})^2} \quad (2)$$

Outra medida de avaliação foi a raiz do erro quadrático médio, RMSE em inglês, que expressa a acurácia dos resultados, apresentando os valores do erro na mesma unidade da variável analisada, [Chai and Draxler 2014], conforme apresentado na Equação 3.

Tabela 1. Análise descritiva e teste de estacionariedade para a série temporal

	Nº Observações	Média	Desvio Padrão	Curtose	Coef. Variação	p_value Dickey-Fuller
Dados Brutos	104	175382.66	49303.68	3.19	0.28	0.000021

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2} \quad (3)$$

3. Resultados e Discussão

A Figura 2, apresenta o gráfico de linhas dos dados da série temporal, nota-se uma sazonalidade nos meses de agosto e nas última semana de dezembro. Essa queda das vendas pode estar diretamente relacionadas ao período em que o empregador pode fechar a fábrica, totalmente ou parcialmente, para as férias dos trabalhadores nos períodos de 15 de Julho a 31 de Agosto, por até 21 dias consecutivos, e/ou durante as férias escolares de Natal, não excedendo 5 dias úteis consecutivos, conforme a cláusula 72, itens b e d do documento [AIMMAP et al. 2010].



Figura 2. Série temporal da quantidade de encomendas semanais

A Tabela 1 apresenta uma breve análise estatística em relação aos dados de venda semanal. Com um total de 104 observações, o coef. de variação indica que as amostras possuem uma dispersão média. O p-value de Dickey-Fuller apresenta um valor menor que 0.05, isso significa que a série temporal é estacionária.

Com a utilização do Prophet, realizou-se o treinamento de alguns modelos em que, alguns parâmetros foram fixados como, o crescimento linear das tendências, e as sazonalidades semanal, mensal e diária, como verdadeira. A lista de feriados definidos, possuía algumas semanas do ano nos meses de julho, agosto e dezembro dos anos de 2016 e 2017.

Os dados apresentados na Tabela 2, onde varia a utilização de alguns parâmetros informados para o treinamento. Vale ressaltar que em ambos dos tipos de sazonalidade, aditiva e multiplicativa, os resultados de coeficiente de determinação e RMSE apresentaram valores melhores quando utilizados com os feriados.

A Figura 3 apresenta as componentes descompostas da série temporal referente ao modelo 1 da Tabela 2. Nota-se que essa série temporal possui aparentemente uma

Tabela 2. Parâmetro do modelo com seus respectivos RMSE e R2

Modelo	Parâmetros	R2	RMSE
01	Sazona.: Aditivo Feriados: Sim	0.757306	34222.431185
02	Sazona.: Multiplicativo Feriados: Sim	0.752305	34573.257638
03	Sazona.: Aditivo Feriados: Não	0.726166	36351.761457
04	Sazona.: Multiplicativo Feriados: Não	0.721847	36637.330079

tendência de crescimento nas encomendas. Observa-se que a tendência, *trend*, teve um crescimento mais acelerado a partir do mês de dezembro de 2016. A componente feriado, *holidays*, apresentou uma diminuição nas encomendas de 30%, como identificado na Figura 2 devido a relação com o período de parada. Em relação a componente anual, *yearly*, observa-se uma queda maior nos meses de julho e agosto, o que já era esperado, devido ser o período em que a fábrica reduz drasticamente as vendas por causa das férias.

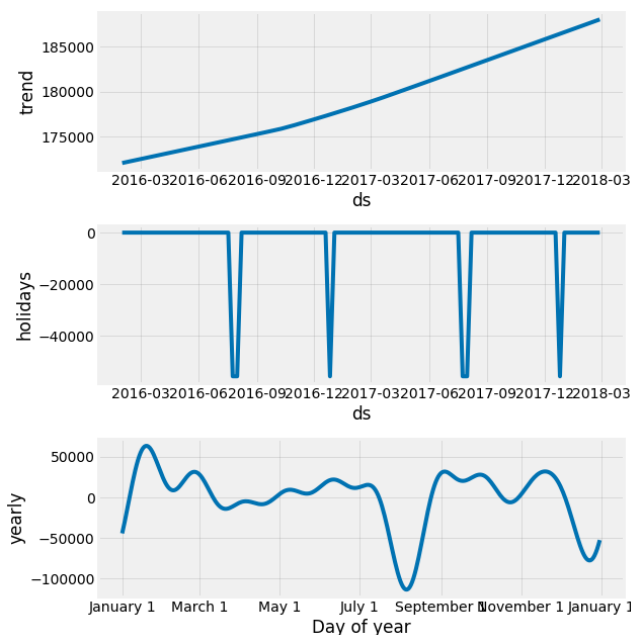


Figura 3. Componentes decompostos da série temporal analisada

4. Conclusão

O agrupamento dos dados de forma semanal realizado nesse trabalho forneceram um melhor entendimento de como a organização funciona, assim foi possível identificar as reduções de encomendas que aconteceram nos períodos de julho, agosto e dezembro.

Esses padrões de sazonalidade poderiam não estar visíveis, se a análise acontecesse separadamente para cada produto. As informações que foram obtidas dessa análise servirão para dar continuidade nos trabalhos futuros.

A utilização da biblioteca Facebook Prophet auxiliou na elaboração e testes dos modelos, além de conseguir obter informação que sobre o modelo e suas principais componentes. A utilização das técnicas clássicas pode ajudar a obter resultados satisfatórios para séries temporais univariáveis. O efeito dos feriados na série temporal fica claro, sendo uma componente interessante a ser explorado. A sazonalidade aditiva foi a que obteve o melhor resultado, conforme a apresentação realizada sobre o modelo e suas explicações.

Referências

- AIMMAP, SINDEL, FETESE, and SITESC (2010). Contrato colectivo de trabalho para a indústria da metalurgia e metalomecânica. http://www.aimmap.pt/fotos/editor2/cct_2010.pdf.
- Chai, T. and Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)? *Geoscientific Model Development Discussions*, 7:1525–1534.
- Costa, P. H. S. and Baidya, T. K. N. (2001). Propriedades estatísticas das séries de retornos das principais ações brasileiras. *Pesquisa Operacional*, 21(1):61–87.
- Damodar, N. et al. (2004). *Basic econometrics*. The Mc-Graw Hill.
- Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Jaynes, E. T. (1980). The minimum entropy production principle. *Annual Review of Physical Chemistry*, 31(1):579–601.
- Mentzer, J. T. and Moon, M. A. (2004). *Sales forecasting management: a demand management approach*. Sage.
- Moon, M. A. et al. (2006). Breaking down barriers to forecast process improvement. *Foresight: The International Journal of Applied Forecasting*, 4:26–30.
- Slack, N., Chambers, S., Johnston, R., et al. (2009). *Administração da produção*, volume 747. Atlas São Paulo.
- Taylor, S. J. and Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1):37–45.

Apêndice C

Tabela dos resultados da previsão realizados pelo modelo RF entre as semanas 5 a 9 do ano de 2018

Tabela C.1: Resultados da previsão realizado pelo modelo da aplicação para a semana do ano 5 a 9 do ano de 2018

Material	MaterialDescription	W+1	Qtd_W+1	W+2	Qtd_W+2	W+3	Qtd_W+3	W+4	Qtd_W+4	W+5	Qtd_W+5
001	Material 001	5	352	6	356	7	352	8	352	9	353
002	Material 002	5	3007	6	3007	7	3007	8	3007	9	3007
003	Material 003	5	247	6	247	7	247	8	247	9	247
004	Material 004	5	31	6	31	7	31	8	31	9	31
005	Material 005	5	128	6	128	7	128	8	128	9	128
006	Material 006	5	5223	6	5223	7	5223	8	5223	9	5223
007	Material 007	5	2173	6	2173	7	2173	8	2173	9	2173
008	Material 008	5	373	6	417	7	417	8	373	9	373
009	Material 009	5	0	6	0	7	0	8	0	9	0
010	Material 010	5	2043	6	2043	7	2043	8	2043	9	2043
011	Material 011	5	2060	6	2060	7	2060	8	2060	9	2060
012	Material 012	5	1395	6	1440	7	1440	8	1440	9	1440
013	Material 013	5	2548	6	2744	7	2822	8	2861	9	2704
014	Material 014	5	2671	6	2716	7	2704	8	2839	9	2844
015	Material 015	5	1105	6	1105	7	1105	8	1105	9	1105
016	Material 016	5	1828	6	1828	7	1828	8	1828	9	1828
017	Material 017	5	1241	6	1241	7	1241	8	1241	9	1241
018	Material 018	5	1692	6	1692	7	1692	8	1692	9	1692
019	Material 019	5	640	6	670	7	590	8	590	9	545
020	Material 020	5	1869	6	1869	7	1869	8	1869	9	1869
021	Material 021	5	235	6	158	7	105	8	230	9	347
022	Material 022	5	1498	6	1426	7	1426	8	1411	9	1531
023	Material 023	5	1480	6	1480	7	1480	8	1480	9	1435
024	Material 024	5	345	6	115	7	460	8	460	9	374
025	Material 025	5	136	6	148	7	196	8	148	9	231
026	Material 026	5	180	6	151	7	180	8	151	9	237
027	Material 027	5	1075	6	1075	7	1075	8	819	9	819
028	Material 028	5	307	6	307	7	307	8	307	9	307
029	Material 029	5	327	6	360	7	396	8	612	9	719
030	Material 030	5	92	6	108	7	71	8	40	9	28
031	Material 031	5	39	6	75	7	44	8	32	9	32
032	Material 032	5	194	6	194	7	194	8	209	9	209
033	Material 033	5	30	6	30	7	30	8	30	9	120
034	Material 034	5	134	6	210	7	210	8	225	9	240
035	Material 035	5	105	6	105	7	90	8	105	9	105
036	Material 036	5	540	6	540	7	540	8	540	9	540
037	Material 037	5	120	6	120	7	120	8	130	9	120
038	Material 038	5	91	6	91	7	91	8	112	9	91

Table C.1 continued from previous page

Material	MaterialDescription	W+1	Qtd_W+1	W+2	Qtd_W+2	W+3	Qtd_W+3	W+4	Qtd_W+4	W+5	Qtd_W+5
039	Material 039	5	436	6	551	7	551	8	592	9	582
040	Material 040	5	15	6	20	7	72	8	10	9	0
041	Material 041	5	104	6	104	7	107	8	114	9	138
042	Material 042	5	1178	6	1178	7	1178	8	1178	9	1178
043	Material 043	5	33	6	35	7	21	8	31	9	33
044	Material 044	5	20	6	31	7	17	8	5	9	5
045	Material 045	5	345	6	316	7	345	8	432	9	489
046	Material 046	5	79	6	19	7	70	8	70	9	50
047	Material 047	5	119	6	77	7	87	8	29	9	49
048	Material 048	5	0	6	0	7	0	8	0	9	0
049	Material 049	5	9	6	9	7	9	8	9	9	9
050	Material 050	5	104	6	89	7	15	8	75	9	89
051	Material 051	5	345	6	374	7	374	8	403	9	374
052	Material 052	5	120	6	104	7	150	8	120	9	135
053	Material 053	5	86	6	66	7	89	8	84	9	84
054	Material 054	5	60	6	30	7	15	8	60	9	50
055	Material 055	5	700	6	879	7	919	8	980	9	1079
056	Material 056	5	180	6	240	7	270	8	345	9	345
057	Material 057	5	0	6	15	7	15	8	15	9	15
058	Material 058	5	30	6	45	7	45	8	15	9	15
059	Material 059	5	45	6	30	7	45	8	104	9	60
060	Material 060	5	225	6	225	7	225	8	225	9	210
061	Material 061	5	104	6	104	7	104	8	135	9	89
062	Material 062	5	823	6	879	7	935	8	1296	9	1344
063	Material 063	5	1279	6	959	7	1216	8	1407	9	1600
064	Material 064	5	968	6	971	7	972	8	967	9	967
065	Material 065	5	830	6	850	7	679	8	680	9	620
066	Material 066	5	100	6	90	7	50	8	90	9	100
067	Material 067	5	127	6	112	7	127	8	127	9	111
068	Material 068	5	7533	6	7631	7	7631	8	7631	9	7679
069	Material 069	5	60	6	60	7	60	8	60	9	60
070	Material 070	5	36	6	0	7	72	8	0	9	12
071	Material 071	5	36	6	24	7	48	8	0	9	24
072	Material 072	5	36	6	0	7	36	8	0	9	12
073	Material 073	5	61	6	61	7	61	8	88	9	88
074	Material 074	5	9	6	9	7	9	8	9	9	9
075	Material 075	5	47	6	47	7	47	8	47	9	47
076	Material 076	5	326	6	326	7	326	8	326	9	326
077	Material 077	5	45	6	45	7	45	8	45	9	45
078	Material 078	5	1240	6	1240	7	1240	8	1240	9	1240

Table C.1 continued from previous page

Material	MaterialDescription	W+1	Qtd_W+1	W+2	Qtd_W+2	W+3	Qtd_W+3	W+4	Qtd_W+4	W+5	Qtd_W+5
079	Material 079	5	1520	6	1520	7	1520	8	1520	9	1520
080	Material 080	5	173	6	173	7	173	8	173	9	173
081	Material 081	5	1056	6	1056	7	1344	8	1152	9	1056
082	Material 082	5	0	6	0	7	0	8	0	9	0
083	Material 083	5	0	6	0	7	0	8	0	9	0
084	Material 084	5	0	6	0	7	0	8	0	9	0
085	Material 085	5	3587	6	3587	7	3587	8	3587	9	3587
086	Material 086	5	19	6	23	7	22	8	20	9	20
087	Material 087	5	1288	6	1100	7	1064	8	988	9	880
088	Material 088	5	0	6	0	7	0	8	0	9	0
089	Material 089	5	0	6	0	7	0	8	0	9	0
090	Material 090	5	882	6	882	7	882	8	882	9	882
091	Material 091	5	129	6	118	7	120	8	119	9	119
092	Material 092	5	377	6	46	7	90	8	108	9	75
093	Material 093	5	17	6	16	7	16	8	16	9	16
094	Material 094	5	1244	6	1244	7	1244	8	1244	9	1244
095	Material 095	5	12	6	12	7	21	8	21	9	4
096	Material 096	5	302	6	220	7	288	8	302	9	271
097	Material 097	5	177	6	175	7	175	8	176	9	177
098	Material 098	5	1080	6	1080	7	1080	8	1082	9	14
099	Material 099	5	10296	6	9328	7	9588	8	9461	9	9026
100	Material 100	5	120	6	120	7	120	8	120	9	120
101	Material 101	5	3878	6	3955	7	3878	8	4031	9	3840
102	Material 102	5	5227	6	6394	7	6250	8	7821	9	5643
103	Material 103	5	945	6	946	7	945	8	945	9	945
104	Material 104	5	25	6	67	7	84	8	84	9	67
105	Material 105	5	1234	6	1192	7	1327	8	1318	9	1297
106	Material 106	5	202	6	202	7	226	8	125	9	193
107	Material 107	5	54	6	45	7	48	8	36	9	30
108	Material 108	5	0	6	54	7	27	8	13	9	54
109	Material 109	5	10	6	10	7	10	8	10	9	10
110	Material 110	5	162	6	160	7	159	8	165	9	159
111	Material 111	5	24	6	24	7	24	8	24	9	24
112	Material 112	5	868	6	806	7	976	8	1004	9	1099
113	Material 113	5	1411	6	1366	7	1440	8	1459	9	1503
114	Material 114	5	652	6	652	7	652	8	652	9	652
115	Material 115	5	7	6	7	7	5	8	5	9	8
116	Material 116	5	0	6	0	7	0	8	0	9	0
117	Material 117	5	901	6	923	7	923	8	890	9	879
118	Material 118	5	197	6	120	7	120	8	109	9	120

Table C.1 continued from previous page

Material	MaterialDescription	W+1	Qtd_W+1	W+2	Qtd_W+2	W+3	Qtd_W+3	W+4	Qtd_W+4	W+5	Qtd_W+5
119	Material 119	5	0	6	0	7	0	8	0	9	0
120	Material 120	5	2736	6	3168	7	2400	8	1584	9	1520
121	Material 121	5	2640	6	2960	7	2480	8	1679	9	1456
122	Material 122	5	0	6	0	7	0	8	76	9	304
123	Material 123	5	120	6	1211	7	1152	8	1020	9	859
124	Material 124	5	48	6	16	7	4	8	0	9	0
125	Material 125	5	0	6	0	7	0	8	0	9	0
126	Material 126	5	28	6	12	7	12	8	12	9	12
127	Material 127	5	1182	6	1182	7	1182	8	1182	9	1182
128	Material 128	5	4	6	4	7	4	8	4	9	4
129	Material 129	5	0	6	0	7	0	8	0	9	0
130	Material 130	5	0	6	0	7	0	8	159	9	159
131	Material 131	5	337	6	374	7	374	8	374	9	324
132	Material 132	5	512	6	512	7	562	8	562	9	562
133	Material 133	5	487	6	525	7	525	8	525	9	525
134	Material 134	5	315	6	315	7	315	8	315	9	315
135	Material 135	5	6829	6	6829	7	6829	8	6829	9	6829
136	Material 136	5	3263	6	3315	7	3250	8	3000	9	2970
137	Material 137	5	2501	6	2723	7	2726	8	2654	9	2622
138	Material 138	5	331	6	331	7	331	8	331	9	331
139	Material 139	5	0	6	0	7	0	8	0	9	0
140	Material 140	5	664	6	973	7	1303	8	887	9	1162
141	Material 141	5	2074	6	1855	7	1855	8	1855	9	1855
142	Material 142	5	114	6	114	7	114	8	114	9	114
143	Material 143	5	979	6	979	7	979	8	979	9	979
144	Material 144	5	1650	6	1650	7	1650	8	1650	9	1650
145	Material 145	5	4004	6	4004	7	4004	8	4004	9	4004
146	Material 146	5	514	6	514	7	514	8	514	9	514
147	Material 147	5	16	6	16	7	16	8	16	9	16
148	Material 148	5	3254	6	3254	7	3254	8	3254	9	3254
149	Material 149	5	57	6	57	7	57	8	57	9	57
150	Material 150	5	115	6	115	7	115	8	115	9	115
151	Material 151	5	70	6	70	7	70	8	70	9	70
152	Material 152	5	244	6	244	7	244	8	244	9	244
153	Material 153	5	331	6	306	7	310	8	307	9	287
154	Material 154	5	0	6	0	7	0	8	0	9	0
155	Material 155	5	108	6	108	7	108	8	108	9	108
156	Material 156	5	340	6	340	7	340	8	340	9	340
157	Material 157	5	0	6	0	7	0	8	0	9	0
158	Material 158	5	0	6	0	7	0	8	0	9	0

Table C.1 continued from previous page

Material	MaterialDescription	W+1	Qtd_W+1	W+2	Qtd_W+2	W+3	Qtd_W+3	W+4	Qtd_W+4	W+5	Qtd_W+5
159	Material 159	5	0	6	0	7	0	8	0	9	0
160	Material 160	5	1	6	1	7	1	8	1	9	1
161	Material 161	5	151	6	151	7	151	8	151	9	151
162	Material 162	5	816	6	864	7	336	8	144	9	623
163	Material 163	5	528	6	528	7	528	8	528	9	528
164	Material 164	5	64	6	80	7	80	8	51	9	80
165	Material 165	5	153	6	192	7	172	8	172	9	192
166	Material 166	5	53	6	73	7	53	8	77	9	53
167	Material 167	5	42	6	42	7	42	8	42	9	42
168	Material 168	5	192	6	200	7	167	8	104	9	184
169	Material 169	5	6	6	9	7	7	8	7	9	11
170	Material 170	5	47	6	40	7	40	8	40	9	39
171	Material 171	5	0	6	0	7	0	8	0	9	0
172	Material 172	5	399	6	227	7	180	8	180	9	319
173	Material 173	5	324	6	180	7	168	8	168	9	388
174	Material 174	5	97	6	97	7	127	8	122	9	115
175	Material 175	5	38	6	38	7	55	8	51	9	25
176	Material 176	5	4	6	9	7	6	8	4	9	5
177	Material 177	5	1332	6	1332	7	1332	8	1477	9	1329
178	Material 178	5	460	6	360	7	352	8	408	9	408
179	Material 179	5	3260	6	2885	7	2922	8	2666	9	2748
180	Material 180	5	36	6	70	7	101	8	94	9	74
181	Material 181	5	639	6	606	7	606	8	512	9	541
182	Material 182	5	258	6	264	7	267	8	258	9	295
183	Material 183	5	137	6	137	7	137	8	137	9	137
184	Material 184	5	872	6	872	7	872	8	872	9	872
185	Material 185	5	32	6	32	7	32	8	32	9	32
186	Material 186	5	63	6	63	7	63	8	63	9	63
187	Material 187	5	341	6	341	7	341	8	341	9	341
188	Material 188	5	1	6	1	7	1	8	1	9	1
189	Material 189	5	2	6	2	7	2	8	2	9	2
190	Material 190	5	3	6	3	7	3	8	3	9	3
191	Material 191	5	2	6	2	7	2	8	2	9	2
192	Material 192	5	3712	6	3712	7	3712	8	3712	9	3712
193	Material 193	5	0	6	0	7	0	8	0	9	0
194	Material 194	5	875	6	875	7	875	8	875	9	875
195	Material 195	5	224	6	224	7	224	8	224	9	224
196	Material 196	5	0	6	0	7	0	8	0	9	0
197	Material 197	5	98	6	98	7	98	8	98	9	98
198	Material 198	5	4	6	4	7	4	8	4	9	4

Table C.1 continued from previous page

Material	MaterialDescription	W+1	Qtd_W+1	W+2	Qtd_W+2	W+3	Qtd_W+3	W+4	Qtd_W+4	W+5	Qtd_W+5
199	Material 199	5	1132	6	1187	7	1176	8	1299	9	1142
200	Material 200	5	24	6	24	7	24	8	28	9	34
201	Material 201	5	0	6	0	7	0	8	0	9	0
202	Material 202	5	300	6	30	7	0	8	120	9	45
203	Material 203	5	1246	6	810	7	1080	8	855	9	855
204	Material 204	5	4267	6	3852	7	3785	8	3572	9	3001
205	Material 205	5	230	6	57	7	144	8	230	9	230
206	Material 206	5	48	6	60	7	35	8	89	9	38
207	Material 207	5	60	6	70	7	79	8	80	9	31
208	Material 208	5	96	6	201	7	158	8	230	9	100
209	Material 209	5	192	6	38	7	192	8	230	9	76
210	Material 210	5	648	6	627	7	680	8	647	9	631
211	Material 211	5	271	6	213	7	642	8	642	9	577
212	Material 212	5	0	6	0	7	0	8	0	9	0
213	Material 213	5	0	6	0	7	0	8	0	9	0
214	Material 214	5	332	6	332	7	332	8	395	9	291
215	Material 215	5	323	6	340	7	343	8	358	9	358
216	Material 216	5	105	6	17	7	0	8	105	9	17
217	Material 217	5	0	6	0	7	0	8	0	9	0
218	Material 218	5	1684	6	1674	7	1674	8	1674	9	1622
219	Material 219	5	1036	6	993	7	993	8	1008	9	993
220	Material 220	5	10080	6	9840	7	10080	8	10080	9	9680
221	Material 221	5	50	6	30	7	46	8	60	9	39
222	Material 222	5	30	6	39	7	27	8	50	9	30
223	Material 223	5	60	6	90	7	79	8	79	9	60
224	Material 224	5	0	6	0	7	0	8	0	9	0
225	Material 225	5	0	6	0	7	0	8	0	9	0
226	Material 226	5	148	6	115	7	154	8	140	9	144
227	Material 227	5	76	6	51	7	128	8	62	9	67
228	Material 228	5	102	6	102	7	131	8	105	9	108
229	Material 229	5	464	6	500	7	500	8	480	9	469
230	Material 230	5	1772	6	1711	7	1834	8	1721	9	1612
231	Material 231	5	639	6	639	7	639	8	639	9	639
232	Material 232	5	1600	6	1600	7	1600	8	1600	9	1600
233	Material 233	5	60	6	60	7	60	8	60	9	60
234	Material 234	5	0	6	0	7	0	8	0	9	0
235	Material 235	5	2284	6	2284	7	2284	8	2284	9	2284
236	Material 236	5	0	6	0	7	0	8	0	9	0
237	Material 237	5	0	6	0	7	0	8	0	9	0
238	Material 238	5	302	6	302	7	302	8	302	9	302

Table C.1. continued from previous page

Material	MaterialDescription	W+1	Qtd_W+1	W+2	Qtd_W+2	W+3	Qtd_W+3	W+4	Qtd_W+4	W+5	Qtd_W+5
239	Material 239	5	288	6	264	7	264	8	167	9	336
240	Material 240	5	2761	6	2761	7	2761	8	2761	9	2761
241	Material 241	5	0	6	63	7	280	8	280	9	251
242	Material 242	5	522	6	542	7	542	8	542	9	640
243	Material 243	5	478	6	478	7	478	8	478	9	448
244	Material 244	5	100	6	100	7	100	8	100	9	100
245	Material 245	5	931	6	900	7	992	8	1000	9	1031
246	Material 246	5	730	6	660	7	779	8	720	9	690
247	Material 247	5	459	6	459	7	459	8	459	9	459
248	Material 248	5	100	6	65	7	90	8	90	9	105
249	Material 249	5	0	6	12	7	21	8	15	9	6
250	Material 250	5	3	6	6	7	9	8	3	9	3
251	Material 251	5	5	6	15	7	28	8	28	9	6
252	Material 252	5	539	6	557	7	543	8	525	9	514
253	Material 253	5	408	6	398	7	388	8	369	9	362
254	Material 254	5	40	6	43	7	38	8	34	9	41
255	Material 255	5	51	6	48	7	47	8	56	9	59
256	Material 256	5	151	6	151	7	151	8	151	9	151
257	Material 257	5	42	6	50	7	39	8	35	9	36
258	Material 258	5	45	6	43	7	34	8	33	9	40
259	Material 259	5	399	6	374	7	374	8	374	9	374
260	Material 260	5	657	6	602	7	626	8	609	9	535
261	Material 261	5	1027	6	977	7	973	8	973	9	864
262	Material 262	5	280	6	264	7	283	8	268	9	204
263	Material 263	5	729	6	717	7	774	8	765	9	666
264	Material 264	5	2440	6	2456	7	2426	8	2426	9	2414
265	Material 265	5	284	6	265	7	289	8	303	9	294
266	Material 266	5	24	6	13	7	13	8	24	9	41
267	Material 267	5	1001	6	1010	7	1036	8	1068	9	1110
268	Material 268	5	745	6	745	7	745	8	745	9	745
269	Material 269	5	14	6	7	7	14	8	7	9	12
270	Material 270	5	14	6	9	7	6	8	24	9	12
271	Material 271	5	0	6	0	7	0	8	0	9	0
272	Material 272	5	276	6	312	7	315	8	294	9	347
273	Material 273	5	768	6	765	7	783	8	795	9	795
274	Material 274	5	552	6	564	7	570	8	564	9	570
275	Material 275	5	9	6	12	7	12	8	18	9	36
276	Material 276	5	147	6	147	7	147	8	147	9	147
277	Material 277	5	39	6	51	7	51	8	37	9	37
278	Material 278	5	26	6	24	7	34	8	26	9	21

Table C.1 continued from previous page

Material	MaterialDescription	W+1	Qtd_W+1	W+2	Qtd_W+2	W+3	Qtd_W+3	W+4	Qtd_W+4	W+5	Qtd_W+5
279	Material 279	5	57	6	58	7	68	8	57	9	59
280	Material 280	5	35	6	35	7	45	8	34	9	36
281	Material 281	5	10	6	10	7	10	8	10	9	10
282	Material 282	5	139	6	139	7	139	8	139	9	139
283	Material 283	5	17	6	16	7	36	8	53	9	21
284	Material 284	5	43	6	41	7	43	8	69	9	65
285	Material 285	5	0	6	3	7	8	8	2	9	19
286	Material 286	5	0	6	0	7	1	8	0	9	0
287	Material 287	5	10	6	10	7	10	8	10	9	10
288	Material 288	5	9	6	9	7	9	8	9	9	10
289	Material 289	5	30	6	35	7	35	8	96	9	48
290	Material 290	5	70	6	52	7	52	8	46	9	80
291	Material 291	5	10	6	12	7	18	8	26	9	26
292	Material 292	5	380	6	380	7	380	8	380	9	380
293	Material 293	5	33	6	30	7	30	8	30	9	30
294	Material 294	5	2460	6	2552	7	2681	8	2702	9	2411
295	Material 295	5	1843	6	1899	7	1759	8	1794	9	2067
296	Material 296	5	25	6	21	7	21	8	21	9	21
297	Material 297	5	612	6	660	7	636	8	612	9	516
298	Material 298	5	0	6	0	7	0	8	0	9	0
Total		5	197184	6	196620	7	198526	8	197919	9	193066