

**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE TECNOLOGIA
COORDENAÇÃO DO CURSO DE ENGENHARIA QUÍMICA**

LUCAS LUCENA DE MORAIS PEREIRA

**SOFT SENSOR EMPREGANDO REDE NEURAL RECORRENTE
LSTM PARA ESTIMAÇÃO DA GRAMATURA NUMA MÁQUINA
DE PAPEL**

**JOÃO PESSOA – PB
2023**

LUCAS LUCENA DE MORAIS PEREIRA

**SOFT SENSOR EMPREGANDO REDE NEURAL RECORRENTE
LSTM PARA ESTIMAÇÃO DA GRAMATURA NUMA MÁQUINA
DE PAPEL**

Trabalho de Conclusão de Curso
apresentado à Coordenação de Engenharia
Química do Centro de Tecnologia da
Universidade Federal da Paraíba, em
cumprimento aos requisitos para obtenção
do título de Bacharel em Engenharia
Química.

Orientador: Prof. Dr. Arioston Araújo de
Moraes Junior

JOÃO PESSOA – PB

2023

**Catalogação na publicação
Seção de Catalogação e Classificação**

P436s Pereira, Lucas Lucena de Moraes.

SOFT SENSOR EMPREGANDO REDE NEURAL RECORRENTE LSTM
PARA ESTIMAÇÃO DA GRAMATURA NUMA MÁQUINA DE PAPEL /
Lucas Lucena de Moraes Pereira. - João Pessoa, 2023.
59 f. : il.

Orientação: Arioston Araujo de Moraes Junior.
TCC (Graduação) - UFPB/CT.

1. Soft sensors. 2. Redes neurais recorrentes. 3.
Gramatura. I. Moraes Junior, Arioston Araujo de. II.
Título.

UFPB/BSCT

CDU 66.01(043.2)

LUCAS LUCENA DE MORAIS PEREIRA

**SOFT SENSOR EMPREGANDO REDE NEURAL RECORRENTE LSTM PARA
ESTIMAÇÃO DA GRAMATURA NUMA MÁQUINA DE PAPEL**

Trabalho de Conclusão de Curso apresentado à Coordenação de Engenharia Química do Centro de Tecnologia da Universidade Federal da Paraíba, em cumprimento aos requisitos para obtenção do título de Bacharel em Engenharia Química.

Aprovado em 09 de junho de 2023.

BANCA EXAMINADORA

Arioston Araújo de Moraes Júnior

Prof. Dr. Arioston Araújo de Moraes Júnior - UFPB

Orientador

Murilo Leite Oliveira

Eng. Murilo Leite Oliveira - Suzano

Examinador

Documento assinado digitalmente



VIVIAN STUMPF MADEIRA

Data: 15/06/2023 08:48:20-0300

Verifique em <https://validar.iti.gov.br>

Engª. Dra. Vivian Stumpf Madeira - UFPB
Examinadora

Wanderson F. A. dos Passos

Eng. Químico Wanderson Felipe Araújo dos Passos - UFPB
Examinador

RESUMO

O trabalho teve como objetivo desenvolver um *soft sensor* para previsão da gramatura do papel. A metodologia adotada consistiu na utilização de uma rede neural recorrente do tipo LSTM (*Long Short-Term Memory*), conhecida por sua capacidade de processar sequências temporais complexas. O *soft sensor* proposto foi treinado e testado em um conjunto de dados composto por diversas variáveis como desempenho da máquina, propriedades de formação da folha e parâmetros do processo. Com base nessas informações, o modelo foi capaz de aprender padrões e relações não lineares presentes nos dados, fornecendo precisão em suas estimativas. Uma análise detalhada dos resultados revelou que a rede neural apresentou desempenho significativamente melhor para baixas gramaturas. Além disso, foi realizada uma comparação entre os valores medidos e os valores previstos pelo *soft sensor*. Observou-se uma forte correlação entre esses dois conjuntos de dados, indicando que as previsões geradas pelo modelo estão consistentes com os valores reais da gramatura. Os resultados obtidos nesse estudo têm implicações significativas para a indústria do papel. O uso de *soft sensors* baseados em redes neurais LSTM pode proporcionar melhorias substanciais no controle de qualidade e no processo de produção, permitindo ajustes em tempo real para garantir o controle das variáveis de produção.

Palavras-chave: *soft sensors*; redes neurais recorrentes; gramatura.

ZUSAMMENFASSUNG

Ziel der Arbeit war die Entwicklung eines Softsensors zur Vorhersage der Papiergrammatur. Die angewandte Methodik bestand in der Verwendung eines rekurrenten neuronalen Netzwerks vom Typ LSTM (Long Short-Term Memory), das für seine Fähigkeit bekannt ist, komplexe Zeitsequenzen zu verarbeiten. Der vorgeschlagene Soft-Sensor wurde an einem Datensatz trainiert und getestet, der aus mehreren Variablen wie Maschinenleistung, Blechverformungseigenschaften und Prozessparametern bestand. Basierend auf diesen Informationen war das Modell in der Lage, Muster und nichtlineare Beziehungen zu lernen, die in den Daten vorhanden sind, und lieferte Genauigkeit in ihren Schätzungen. Eine detaillierte Analyse der Ergebnisse ergab, dass das neuronale Netzwerk bei niedrigen Grammaturen eine deutlich bessere Leistung zeigte. Zusätzlich wurde ein Vergleich zwischen den Messwerten und den vom Soft-Sensor vorhergesagten Werten durchgeführt. Es gab eine starke Korrelation zwischen diesen beiden Datensätzen, was darauf hindeutet, dass die vom Modell generierten Vorhersagen mit den tatsächlichen Grammaturwerten übereinstimmen. Die Ergebnisse dieser Studie haben erhebliche Auswirkungen auf die Papierindustrie. Der Einsatz von Soft-Sensoren auf Basis neuronaler LSTM-Netzwerke kann die Qualitätskontrolle und den Produktionsprozess erheblich verbessern und Echtzeitanpassungen ermöglichen, um die Kontrolle der Produktionsvariablen sicherzustellen.

Schlüsselwörter: Soft-Sensors; rekurrente neuronale Netze; Grammatur.

SUMÁRIO

1	INTRODUÇÃO.....	7
1.1	OBJETIVO GERAL.....	9
1.1	OBJETIVOS ESPECÍFICOS.....	9
2	FUNDAMENTAÇÃO TEÓRICA	9
2.1	PROCESSO DE FABRICAÇÃO DO PAPEL	9
2.2	MÁQUINA DE PAPEL.....	13
2.2.1	Preparação da massa	13
2.2.2	Círculo de aproximação	16
2.2.3	Formação do papel	17
2.2.3.1	Caixa de entrada.....	18
2.2.3.2	Mesa formadora	19
2.2.3.3	Prensagem	20
2.2.3.4	Secagem	21
2.2.3.5	Colagem superficial	23
2.2.3.6	Calandragem	24
2.2.3.7	Enrolamento	24
2.3	INTRODUÇÃO A <i>SOFT SENSORS</i>	25
2.3.1	Coleta e tratamento de dados	27
2.3.1.1	Tempo de amostragem	27
2.3.1.2	Dados ausentes	28
2.3.1.3	<i>Outliers</i>	28
2.3.2	Seleção de variáveis de entrada	29
2.3.3	Escolha e treinamento do modelo.....	31
2.3.4	Validação do modelo	31
2.3.5	Manutenção do <i>soft sensors</i>	32
2.4	REDES NEURAIS ARTIFICIAIS	33
2.4.1	Tipos de redes neurais artificiais	34
2.4.2	<i>Long Short-Term Memory</i>	35
3	METODOLOGIA.....	36
3.1	HISTÓRICO E COLETA DE DADOS	36
3.2	MÉTODO DE SELEÇÃO DE VARIÁVEIS DE ENTRADA	38
3.3	ESTRUTURA DO MODELO LSTM PROPOSTO	40

4	RESULTADOS E DISCUSSÕES	42
5	CONCLUSÕES.....	50
	REFERÊNCIAS BIBLIOGRÁFICAS.....	51
	APÊNDICE A – TOPOLOGIA DA REDE NEURAL ARTIFICIAL PROPOSTA NO TRABALHO	55
	ANEXO A – MODELO MATEMÁTICO DA RNN LSTM EXTRAÍDO DE LIU & GONG, 2019.....	56

1 INTRODUÇÃO

Com a revolução industrial, bem como com o surgimento dos meios de comunicação impressos (livros, jornais e revistas), o consumo de papel aumentou de forma significativa, fazendo com que houvesse um grande avanço na área (SHREVE; BRINK JR, 1980). Assim, importantes tecnologias surgiram a partir do século XVIII, as quais permitiram a fabricação de papel a partir de madeiras, aumentando significativamente a capacidade de produção devido à maior disponibilidade dessas matérias-primas dos que as tradicionais fibras utilizadas (TEIXEIRA *et al*, 2017).

As máquinas utilizadas na produção de papel antigamente geralmente exigiam a intervenção direta dos trabalhadores para realizar as operações. Por exemplo, era comum os operadores ajustarem manualmente as configurações das máquinas, controlarem o fluxo de matérias-primas e monitorarem visualmente o processo para garantir a qualidade do produto final, e desta forma, representavam riscos significativos à segurança dos trabalhadores. Os operadores precisavam interagir diretamente com peças e partes móveis das máquinas, expondo-se ao perigo de ferimentos graves, como cortes, esmagamentos ou aprisionamentos.

Devido a necessidade de aumentar a fabricação decorrente do aumento da demanda, se fez necessário desenvolver novas matérias primas e novas técnicas para fabricação do papel, assim, introduziu-se a madeira existente nos troncos das árvores e foram incorporadas as máquinas que permitiam a produção de forma contínua, e não mais em folhas separadas (PESCIO, 2016).

Com o avanço da tecnologia e a automação na indústria de papel, houve uma redução significativa da necessidade de contato direto homem-máquina. As máquinas modernas são projetadas com recursos de segurança aprimorados, automação de processos e sistemas de controle mais avançados, minimizando a exposição dos trabalhadores a situações perigosas e melhorando a segurança geral no ambiente de trabalho. Ao longo das últimas duas décadas, essa indústria passou por transformações significativas, impulsionadas por avanços na tecnologia digital, preocupações ambientais crescentes, aumento dos custos de energia e mão de obra, além de uma busca cada vez maior por objetivos de sustentabilidade.

A modernização do processo de produção de papel tem revolucionado a forma como as operações são conduzidas na indústria de celulose. Alguns aspectos-chave da modernização incluem:

- a) **Automação e digitalização:** Utilização de sistemas automatizados e tecnologias digitais para controlar e monitorar os processos de produção, minimizando erros e otimizando a eficiência.
- b) **Otimização de recursos:** Maximizar o uso de matérias-primas, água, energia e outros recursos, reduzindo o desperdício e aumentando a produção.
- c) **Utilização de tecnologias mais eficientes:** A implementação de tecnologias avançadas, como sistemas de recuperação de calor, cogeração de energia e tecnologias de branqueamento de polpa mais eficientes, contribui para reduzir o consumo de energia e minimizar os impactos ambientais.
- d) **Melhoria da qualidade do produto:** A modernização dos processos permite a produção de papel com características específicas, como maior resistência, melhor acabamento e maior uniformidade.
- e) **Sustentabilidade ambiental:** A modernização do setor de papel busca reduzir o impacto ambiental por meio da adoção de práticas sustentáveis, como o uso de fibras recicladas, redução do consumo de água e energia, e tratamento adequado dos efluentes.

A indústria de papel atual se beneficia de avanços tecnológicos significativos, especialmente na área da automação. Por meio da utilização de controladores lógicos programáveis e sistemas supervisionados, busca-se alcançar a otimização do processo, além de permitir um monitoramento contínuo em todas as etapas da cadeia de produção.

Segundo Pescio (2016), a integração dos sistemas de automação e o uso de softwares avançados facilitam o gerenciamento a partir de uma única sala de operação, resultando numa melhor comunicação, e consequentemente, numa gestão mais eficaz das execuções. Além de agrupar as operações no mesmo ambiente, também é possível disponibilizar as informações de cada etapa do processo produtivo com respostas rápidas e precisas.

A integração do sistema de automação, além de ser utilizada como ferramenta que permite o controle e a supervisão do processo produtivo, também estabelece uma harmonia funcional dos sistemas de produção, podendo ser utilizada na organização do trabalho dos operadores e da manutenção, auxiliar na melhoria da qualidade do produto final e ainda disponibilizar os dados do processo produtivo e dos equipamentos para o nível gerencial, contribuindo nas tomadas de decisões. Esta integração consiste basicamente em alocar componentes e sistemas de maneira conjunta levando a um estado de sinergia entre eles (PESCIO, 2016).

Dessa forma, a indústria de celulose e papel continua evoluindo para atender aos desafios e demandas do século XXI, aproveitando as oportunidades trazidas pela tecnologia digital, adotando práticas ambientalmente responsáveis e buscando um equilíbrio entre eficiência operacional e produtividade.

1.1 OBJETIVO GERAL

O objetivo geral deste trabalho é desenvolver e implementar um *soft sensor* baseado em redes neurais *LSTM* (*Long Short-Term Memory*) para a predição da gramatura do papel usando dados de um processo industrial.

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste trabalho são:

1. Realizar um levantamento bibliográfico sobre os conceitos e técnicas de *soft sensors*, redes neurais *LSTM* e sua aplicação em processos industriais.
2. Coletar e pré-processar os dados, realizando a limpeza, normalização e seleção das variáveis mais relevantes para o aprendizado de máquina através de análises de sensibilidade para identificar a influência no modelo e seu impacto na predição da gramatura.
3. Projetar e treinar uma rede neural *LSTM* utilizando os dados disponíveis, ajustando os parâmetros do modelo para obter o melhor desempenho, avaliando por meio de métricas de qualidade, como erro médio, erro quadrático médio e coeficiente de correlação.
4. Comparar os resultados da predição da gramatura obtidos pelo *soft sensor* com os valores reais medidos para verificar a acurácia da rede neural.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 PROCESSO DE FABRICAÇÃO DO PAPEL

A principal matéria-prima para a produção de papel é a celulose, um polímero de cadeia longa, muito abundante na madeira, podendo também ser encontrado em folhosas e em frutos, como o algodão (D'ALMEIDA *et al*, 2013) (SANTOS *et al*, 2001). Além da

celulose, outros compostos podem ser encontrados na madeira, tais como as hemiceluloses, a lignina e os extrativos da madeira. Enquanto a celulose representa 40-50% da composição química da madeira, as hemiceluloses representam 25-35%, a lignina, 10-30% e os extrativos, 0,5-5% (TEIXEIRA *et al*, 2017).

As hemiceluloses são polissacarídeos ramificados formados por pelo menos duas unidades de açúcares, como as hexoses e pentoses e as oxihexoses e ácidos urônicos. Têm baixo grau de polimerização, são materiais sólidos brancos e têm natureza cristalina, características que conferem às hemiceluloses a propriedade de absorverem água com facilidade. Além disso, apresenta estrutura fibrosa, a qual está associada às fibras da celulose (D'ALMEIDA *et al*, 2013).

A presença da lignina na madeira não tem valor para a indústria de celulose e papel, de forma que o principal objetivo para a obtenção da polpa de celulose é a solubilização da lignina que ocorre no processo denominado de polpação (SHREVE; BRINK JR, 1980) (D'ALMEIDA *et al*, 2013).

Os extrativos são compostos de naturezas diversas, apresentam alta massa molecular e são solúveis em solventes orgânicos ou em água. Esses compostos conferem cor, odor, sabor e resistência à degradação de algumas espécies, de forma que variam bastante em cada espécie de madeira. Normalmente, os extrativos encontrados na madeira são os terpenos e seus derivados, os triglicerídeos (óleos e gorduras e seus derivados) e compostos fenólicos, podendo também ser encontrados aminoácidos, açúcares solúveis e alcaloides (D'ALMEIDA *et al*, 2013).

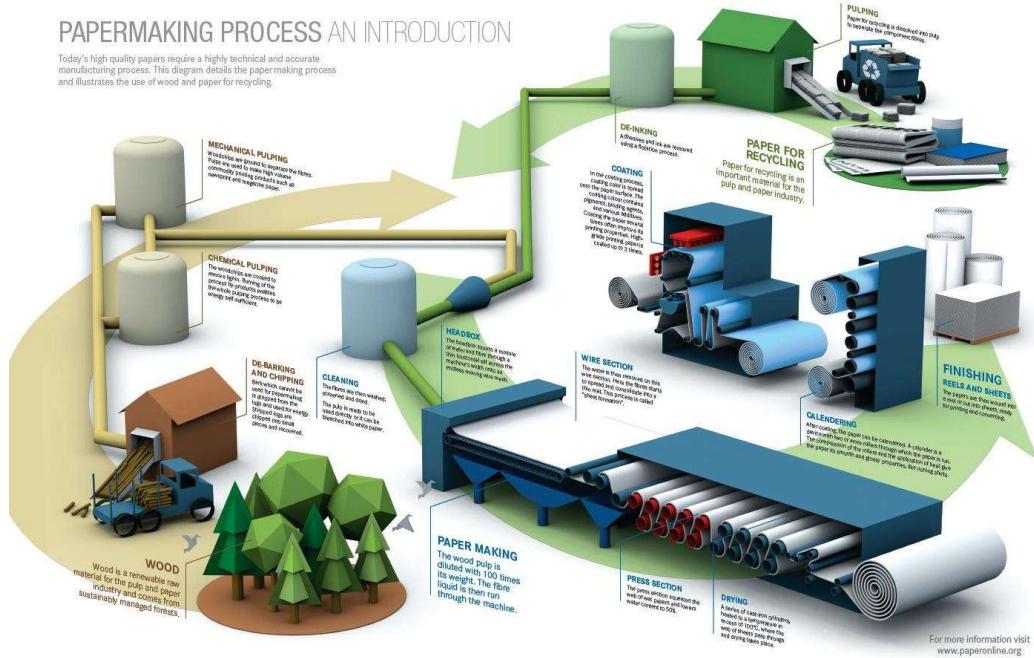
A indústria de celulose e papel realiza processos químicos e operações unitárias com o objetivo de obter o sólido final composto por celulose e hemiceluloses, solubilizando a lignina, podendo ser englobados nos seguintes processos: (1) Preparação da madeira; (2) Extração da celulose ou polpação; (3) Etapa de branqueamento; (4) Refinação; e (5) Secagem e papel acabado (TEIXEIRA *et al*, 2017).

O processo de produção de papel se inicia pela preparação da madeira em condições ideais para o processo de polpação. Inicia-se pela lavagem das toras de madeira, a fim de remover areia ou terra acumulada durante a manipulação dos troncos na floresta ou no pátio da indústria. Após a lavagem, segue-se para a remoção da casca da madeira, também denominado de descascamento (SHREVE; BRINK JR, 1980) (D'ALMEIDA *et al*, 2013).

Após o descascamento, as toras são encaminhadas para a picagem, sendo cortadas em pedaços denominados de cavacos, cuja dimensão deve estar de acordo com o

processamento seguinte, a polpação, visto que a sua espessura interfere diretamente no consumo dos reagentes e consequente rendimento da polpa celulósica obtida. Para isso, os cavacos são classificados por peneiramento, onde os de tamanho médio são encaminhados para o processo de polpação, os maiores retornam para a picagem e os menores são encaminhados para a queima em caldeira (D'ALMEIDA *et al*, 2013).

Figura 2.1 – Exemplo de processo de produção de papel



Fonte: ALVES; FIORIO, 2014

Como mencionado anteriormente, a madeira é uma matéria-prima constituída por fibras de celulose que se encontram misturadas com hemicelulose, lignina e extractos. Para a formação da folha do papel, são usadas essencialmente as fibras de celulose, sendo necessário o uso de energia mecânica ou química para separá-las dos demais constituintes e formar uma massa celulósica denominada polpa (SHREVE; BRINK JR, 1980) (D'ALMEIDA *et al*, 2013).

O processo de polpação química é o mais utilizado na indústria de celulose e papel, pois promove a remoção total da lignina (deslignificação) das fibras de celulose e são obtidas fibras mais longas. No processo de polpação alcalina, também denominado de *kraft* – forte, em alemão – os cavacos de madeira são digeridos em solução química, sob uma pressão de 6,8 a 9,2 atm, temperatura de 170 a 177 °C, durante 2 a 5 h, podendo ser aplicado a uma grande variedade de espécies de madeira (SHREVE; BRINK JR, 1980).

Nesse processo, ocorre a deslignificação sem a degradação ou remoção da celulose, o que torna a polpa obtida de elevada resistência. A solução química utilizada no processo de digestão, denominada "licor de cozimento" ou "licor branco", consiste numa solução aquosa de hidróxido de sódio e sulfeto de sódio (D'ALMEIDA *et al*, 2013).

Os íons hidrossulfeto reagem seletivamente com a lignina, promovendo a quebra de sua cadeia, e preservando as moléculas de celulose. Já os íons hidróxido apresentam seletividade dependente da temperatura, promovendo a hidrólise da lignina, das hemiceluloses e da celulose, sendo que a reação com esta última ocorre em menor escala. Assim, a temperatura e a concentração do licor branco devem ser ajustadas de forma a promover a obtenção de um cozimento ou polpação ideal, retirando uma maior quantidade de lignina sem degradar as fibras de celulose da polpa (D'ALMEIDA *et al*, 2013).

Após o processo de polpação, a polpa celulósica obtida possui coloração marrom devido a parte de lignina dissolvida estar ainda impregnada na celulose. Assim, ela pode ser submetida a um processo de branqueamento para remoção da lignina residual, o qual consiste no uso de agentes branqueadores, como o hidróxido de sódio, cloro e seus compostos (hipoclorito e dióxido de cloro) e ozônio. Esse processo de branqueamento pode ser considerado como uma continuação da deslignificação, ou seja, ele completa a remoção da lignina iniciada no processo de polpação (SHREVE; BRINK JR, 1980) (PIOTTO, 2003).

Segundo Pescio (2016), é interessante dizer que, o processo fundamental desenvolvido pelos chineses para fabricar o papel são relativamente os mesmos utilizados nos dias atuais, porém, a fabricação de papel moderna passou a ser uma instalação industrial de grande complexidade, com processos e máquinas contínuas de alta tecnologia e elevada capacidade de produção, podendo ser dividida em: (a) fábrica de papel integrada a fábrica de celulose e (b) fábrica de papel não integrada a fábrica de celulose.

A fabricação integrada de celulose e papel ocorre em uma unidade de manufatura, em que a madeira é recebida na forma de cavacos e passa por uma série de etapas para preparar a pasta de celulose.

Essa pasta, ainda em estado úmido, é então transportada por meio de tubulações até a máquina de papel, onde é transformada em uma folha contínua. Esse processo contínuo e conectado permite que toda a fabricação, desde a obtenção da celulose até a produção do papel final, seja realizada de forma eficiente e coordenada em uma única instalação.

2.2 MÁQUINA DE PAPEL

Para ser classificado como papel, as folhas finas devem ser feitas a partir de fibras que foram maceradas até que cada filamento individual seja uma unidade separada, estas fibras misturadas com água passam por uma peneira do tipo tela, onde então são separadas da água pela drenagem através das pequenas aberturas na tela, formando assim uma folha fina de fibras entrelaçadas sobre a superfície da tela. Esta fina camada de fibra entrelaçada é o papel (HUNTER, 1978).

Figura 2.2 – Estrutura da máquina de papel da *VOITH*



Fonte: VOITH, 2017

2.2.1 Preparação da massa

Nesta etapa do processo, a matéria prima fibrosa é submetida a tratamentos mecânicos, físico-químico, à adição de produtos químicos e outros componentes fibrosos ou não fibrosos necessários à fabricação do papel, com características e propriedades específicas para cada tipo de papel a ser fabricado (ROBUSTI *et al*, 2014) (JUDASZ, 2009).

O preparo da massa é o conjunto de ações que antecedem a formação do papel e visa propiciar às fibras todas as condições para satisfazer os parâmetros de qualidade do

produto final, portanto, não existe um sistema de preparo de massa padrão, as diferentes fases do processo e os tipos de equipamentos utilizados, são determinadas para atuar de diferentes formas, de acordo com o tipo de matéria-prima e com o tipo do papel a ser fabricado (PESCIO, 2016).

Existem várias formas de introduzir fibras na preparação de massa. em fábricas integradas à fabricação de celulose, isso pode ser feito por tubulações, onde as fibras são transportadas como suspensão com consistência de 3% a 6%. Outra opção é usar folhas ou rolos de fibras com umidade de aproximadamente 10% provenientes de uma secadora de celulose. Também é possível utilizar fibras com umidade de cerca de 60% de uma desaguadora. A escolha do método depende das necessidades e características do processo de fabricação.

Exceto no caso das fábricas integradas, existe a necessidade de formar uma suspensão de fibras em água com a consistência adequada para poder utilizá-la nas etapas posteriores do processo. Isto é obtido através da etapa de desagregação. O desagregador, também chamado de “*hydropulper*” ou simplesmente “*pulper*”, é um equipamento semelhante a um liquidificador gigante, cuja estrutura corresponde a um tanque de aço, de ferro fundido, de concreto, ou mesmo, com azulejos. O formato do tanque e seu tamanho são determinados pela capacidade de produção de papel requerida e da consistência de operação (CAMPOS, 2011).

A qualidade do papel é amplamente influenciada pelo nível de limpeza da massa utilizada, por isso a depuração é um processo que tem como objetivo remover sujeiras e outros corpos estranhos que são indesejáveis para a aparência e finalidade da folha de papel.

As impurezas podem ser provenientes de diversas fontes, como matéria-prima fibrosa, papéis reciclados, impurezas introduzidas durante o transporte, produtos químicos, equipamentos, água e processo de fabricação, entre outros. É fundamental garantir um processo de depuração eficiente para assegurar a obtenção de um papel de qualidade.

Como relatado por Campos (2011), a eliminação das impurezas é necessária tanto do ponto de vista do papel acabado como do próprio funcionamento do processo. Com relação ao papel, a presença de impurezas afeta as características mecânicas (a impureza é um ponto débil) e ao aspecto exterior (alvura, presença de pastilhas). Com relação à fabricação, a presença de impurezas volumosas ou gelatinosas pode causar perfurações e quebras no papel na saída da máquina, na seção de prensas e na seção de secagem. Os

motivos pelo quais as impurezas são removidas nos processos de fabricação de papel, bem como a eficiência de depuração, dependem de cada processo.

Existem dois sistemas de depuração amplamente conhecidos na fabricação de celulose e papel: o peneiramento (*screen*) e a centrifugação (*cleaner*). No caso do peneiramento, a separação dos componentes da massa é baseada principalmente no tamanho e formato, enquanto que no caso dos depuradores centrífugos, a separação é feita com base na massa específica dos componentes.

Os separadores centrífugos têm como objetivo separar de forma contínua materiais estranhos ao papel presente na suspensão da massa. Esses materiais são partículas de sujeira de elevada massa específica, ou seja, impurezas que são especificamente mais pesadas do que a água. Tratam-se, sobretudo, de grãos de areia, lascas finas de vidro, metais e outros (COMMISSION, 2001). Em seguida a massa é bombeada para o tanque de massa refinada passando pelo processo de refinação, uma das etapas de suma importância para proporcionar as características e qualidade da folha de papel (PESCIO, 2016).

O processo de refinação é o tratamento mecânico que modifica irreversivelmente as características das fibras, tende a aumentar à flexibilidade, a fibrilação interna e externa, a formação de finos e o encurtamento das fibras, concedendo ao material fibroso uma estrutura adequada em função das características desejadas no papel a ser fabricado (ROBUSTI et al., 2014) (PHILIPP; D'ALMEIDA, 1988)

Estes fatores tornam as fibras ramificadas que se entrelaçam com mais facilidade umas às outras, aumentando as forças de ligação entre as fibras que vão compor a folha de papel, com isso, o produto final ganha resistência à tração, perde porosidade e tem um perfil mais homogêneo, todas estas características em função do grau de refinação (SÉKULA, 2011) (MENDES, 2006).

Os refinadores de discos duplos (“*Double-Disc Refiner*”) oferecem alta eficiência e bom potencial para tratamento por causa de duas zonas no refinador. De um modo geral, esses equipamentos possuem dois discos, um estator e um rotor. A massa é alimentada sob pressão através de um orifício que está no centro do disco estator e, em seguida, é obrigada a passar entre os discos, onde é refinada (CAMPOS, 2011).

Após o processo de desagregação (da massa refugada dentro do processo de fabricação de papel), limpeza e refinação, a massa refinada será bombeada para o processo que antecede a formação do papel, também conhecido como circuito de aproximação ou *approach flow* (PESCIO, 2016).

2.2.2 Circuito de aproximação

O circuito de aproximação corresponde ao elo entre a área da preparação de massa e a unidade de formação da máquina de papel. Sua função básica é o transporte da suspensão até a máquina, proporcionando um fluxo estável, homogêneo e limpo. Dele, fazem parte os sistemas de alimentação de massa grossa (caixa de nível, válvula de gramatura, etc.), e de diluição, de “centricleaners”, de peneiras pressurizadas e de controle de fluxo. Trata-se de um conjunto essencial para a fabricação de um produto de boa qualidade. Este conglomerado de equipamentos estáticos e dinâmicos de vários fornecedores deve funcionar junto em todas as faixas de gramatura e velocidades da máquina de papel (CAMPOS, 2011).

Parâmetros tais como perfil longitudinal de gramatura, formação da folha, ausência de contaminantes no papel podem ser influenciados e controlados através do circuito de aproximação. Fenômenos tais como flutuações de consistência e/ou de fluxo, pulsações hidráulicas, vibrações mecânicas, admissão de ar, velocidades inadequadas no interior das tubulações e acabamento inapropriado das superfícies internas das mesmas, influenciarão negativamente os parâmetros acima. Portanto, um circuito de aproximação bem projetado deverá oferecer condições para que tais fenômenos não ocorram ou que sejam minimizados a níveis não prejudiciais à qualidade desejada (CAMPOS, 2011).

Um sistema corretamente projetado deve ter:

- a) **Estabilidade de fluxo e controle satisfatório:** o sistema deve ser capaz de manter um fluxo estável de massa, garantindo uma alimentação consistente para o processo de depuração. Além disso, o controle do sistema deve ser eficiente, permitindo ajustes adequados conforme necessário.
- b) **Boa mistura e dispersão de fibras:** O sistema deve promover uma mistura homogênea das fibras na massa, assegurando uma distribuição uniforme dos componentes e evitando a formação de aglomerados. Isso contribui para a eficiência da separação e remoção de impurezas.
- c) **Sistema de limpeza e facilidade para lavagem:** O sistema deve ser projetado com recursos adequados de limpeza, permitindo a remoção eficiente de impurezas retidas nas telas, cilindros ou outros elementos do equipamento. Além disso, a facilidade de lavagem é importante para manter a higiene do sistema e evitar o acúmulo de resíduos ao longo do tempo.

A folha de papel com maior qualidade e as melhorias havidas nas máquinas de papel, resultaram em sistemas de aproximação cada vez mais complexos. Caixas de entrada hidráulicas, alta velocidade nas máquinas e melhoria nas características da folha exigem que o sistema de aproximação inclua múltiplos estágios de separadores centrífugos, depuração e melhor controle do ar contido no fluxo de massa (CAMPOS, 2011).

A chamada depuração de cabeça de máquina, que faz parte do “*approach flow*”, tem por objetivo proporcionar ao formador uma composição livre de impurezas que possa afetar a qualidade final do papel e a produtividade da máquina (as impurezas podem ser as causas de quebras da folha de papel durante a produção). Desde o ponto de vista de impurezas nas matérias prima, estas devem ser eliminadas eficientemente nos circuitos de preparação de massa, restando, em todos os casos, a depuração de cabeça de máquina como um mero “filtro” prévio ao formador (CAMPOS, 2011).

Após a preparação e diluição da massa o próximo passo é a formação do papel onde a suspensão fibrosa é entregue a máquina de papel através da caixa de entrada que faz a conexão entre o circuito de aproximação e a máquina de papel (PESCIO, 2016).

2.2.3 Formação do papel

O processo definitivo da formação da folha de papel tem início quando a suspensão fibrosa proveniente do preparo de massa é entregue ao primeiro componente da máquina de papel, a caixa de entrada. A massa e os aditivos são misturados nesta suspensão com concentração menor que 1,5%, a qual é distribuída sobre uma tela formadora para desaguamento. Após o desaguamento no formador, a folha passa por prensas para retirada da maior quantidade de água possível antes de seguir pelos cilindros secadores, em seguida para melhorar o acabamento superficial a folha segue para os cilindros de calandragem, e uma vez o papel seco é enrolado continuamente em bobinas chamadas de jumbo (PESCIO, 2016).

A máquina de papel é projetada com dois processos principais: a parte úmida e a parte seca. A parte úmida é responsável por formar a folha de papel e remover a água, consistindo na caixa de entrada, formador e prensas. Por outro lado, a parte seca tem o objetivo de aumentar o teor seco do papel, promover o acabamento superficial e realizar o enrolamento contínuo. Essa etapa inclui os processos de secagem, calandragem e enrolamento.

2.2.3.1 Caixa de entrada

Dentro do processo de fabricação de papel, a caixa de entrada ocupa uma posição de extrema importância. Representa o elo de união entre a parte constante (“*approach flow*”) da instalação de preparação de massa e a máquina de papel (CAMPOS, 2011).

A função da caixa de entrada é distribuir a massa fibrosa em suspensão, ao longo de toda a largura da máquina, a um fluxo com volume constante, quanto ao tempo e ao ponto de incidência na zona de formação da folha, com concentração uniforme do material fibroso e espessura adequada da suspensão (LINDSTRÖM, 2008) (HENTINEN, 2010).

A alimentação da suspensão fibrosa para a máquina de papel deve ser realizada de maneira mais uniforme possível, com fluxo perfeitamente regular e homogêneo sobre toda a largura, com as fibras dispersadas uniformemente, sem agrupamento em flocos, torvelinhos ou correntezas; o fluxo da suspensão deverá ter uma velocidade constante e compatível com a velocidade da tela. A utilização de baixas consistências nos permite obter esta homogeneização e distribuição uniformemente dispersas na suspensão. Com altas consistências, existe a tendência de formar flocos de fibras que afetam a formação da folha (CAMPOS, 2011).

A caixa de entrada recebe a suspensão fibrosa, proveniente da preparação da massa e a distribui na tela formadora, em forma de jato, para isso, a máquina é operada com sistema de controle, onde possibilita que a relação entre a velocidade do jato e a velocidade da tela formadora da máquina seja determinada pelo operador em função do papel produzido. A relação jato/tela é a diferença da velocidade entre o jato da suspensão fibrosa que saída da caixa de entrada e a velocidade da tela formadora (MORAES; PESCIOL, 2010).

O ponto básico que define a eficiência da caixa de entrada é a qualidade do jato de massa que a mesma proporciona. As principais características de qualidade do jato são as seguintes: jato transversal uniforme de massa, mistura homogênea, fluxo livre de pulsações, boa estabilidade, atenuação de pulsos, turbulência controlada, eficiente geometria do lábio, livre de fluxos transversais, perfil uniforme de gramatura e adequada relação jato-tela (CAMPOS, 2011).

A maioria das máquinas trabalha com atraso do jato. Assim, o movimento relativo entre a tela e a massa diminui o número de flocos pelo cisalhamento mecânico, contribuindo para uma boa formação e alinhamento das fibras na direção de máquina.

Quando as velocidades do jato e da tela são iguais, dizemos que estamos trabalhando no teórico. Na condição em que a velocidade do jato é maior que a velocidade da tela, uma fração da água contida no jato volta para trás com grande probabilidade de formar flocos de fibras afetando diretamente na formação da folha (CAMPOS, 2011).

2.2.3.2 Mesa formadora

A seção de formação utilizada na máquina consiste de uma tela sem fim que escorrega sobre uma série de elementos desaguadores proporcionando um desaguamento gradativo, contínuo e suave, suportada em uma estrutura física e adequada para tal operação, esta é a primeira etapa da formação da folha de papel (PESCIO, 2016).

A formação da folha ocorre devido à suspensão das fibras de celulose que são colocadas ao longo da tela formadora pelo jato proveniente da caixa de entrada com consistência variando em média de 0,1% a 1,5% de fibras, enquanto percorre sobre a tela formadora, parte da água escoa por forças gravitacionais e parte é sugada pela ação do vácuo das caixas de sucção, durante este processo as fibras começam a aderir umas às outras formando uma espécie de tecido com fios muito pequenos e trançados entre si, dando início a formação da folha de papel. Quando a folha de papel deixa a seção de formação o índice de sólidos seco é de aproximadamente 20% (PHILIPP; D'ALMEIDA, 1988) (NUNES, 2007).

Para dispor de uma boa qualidade na impressão, principalmente quando se imprime grandes áreas, a distribuição das fibras deve ser o mais uniforme possível. Se existe em determinadas zonas agrupamentos de fibras de distinta espessura ou, simplesmente, uma maior presença destas fibras, a resposta de muitas outras características do papel opacidade, estabilidade dimensional, resistência, etc., serão claramente distintas naquelas áreas. Devido a maior acumulação de fibras, haverá uma pressão muito maior exercida pela calandra naquela zona causando uma série de inconvenientes como níveis de aspereza diferenciados (CAMPOS, 2011).

Porém, nem tudo são fibras na formação do papel. Os materiais de carga e os aditivos também devem ser considerados e devem ser escolhidos adequadamente para aumentar a opacidade, a brancura ou textura; para reduzir a absorvência da tinta, para melhorar as qualidades mecânicas, etc. À medida que se aumenta a porcentagem destes materiais no papel se vai reduzindo o tamanho dos espaços vazios e dos capilares em seu interior (CAMPOS, 2011).

Em máquinas com formador tipo *Duoformer*, pode-se conseguir um desaguamento praticamente simétrico nas duas faces do papel, isto devido a ação de sucção das caixas formadoras inferior e superior, e pela pressão das réguas sobrepostas da caixa de formação inferior e da caixa de sucção superior, as quais provocam forças de cisalhamento na suspensão que se encontra no sanduíche tela-suspensão-tela, minimizando assim a formação de flocos de fibras. Em decorrência do desaguamento bilateral, obtém-se papéis com o mesmo teor de fibras em ambas as faces, o que é uma condição para a igualdade entre as faces do papel, caracterizando uma melhor formação e consequentemente um papel com melhores características (PESCIO, 2016).

2.2.3.3 Prensagem

A prensagem úmida era originalmente baseada na compressão mecânica da folha de papel entre dois rolos sólidos. Esta compressão faz a água escoar a partir do “nip”, isto é, da linha de menos distância entre os rolos, através do rolo inferior. Em máquinas de baixa velocidade, o uso de rolos lisos e sólidos era um modo eficiente de remoção de água, mas não adequado em máquinas com velocidades altas, devido à necessidade de se retirar maior quantidade de água em menor tempo que a capacidade que os filtros possam comportar, causando assim o esmagamento da folha (CAMPOS, 2011).

Figura 2.3 – Exemplo de prensa de papel de uma máquina *VOITH*



Fonte: VOITH, 2017

Nas máquinas atuais, a prensagem é feita com rolos de sucção ou rolos ventilados que possuem capacidade de remoção de água muito maior que as prensas de rolos lisos. A remoção da água é efetuada pela ação combinada da compressão mecânica e da sucção, aplicada através da estrutura porosa do rolo inferior (CAMPOS, 2011).

A seção das prensas é constituída por rolos de pressão, caixas de vácuo, feltros e rolo *pick-up* (responsável por retirar e transferir a folha da seção de formação para a seção de prensas). Esta seção tem como objetivo remover a água do papel através da compressão mecânica, entre os rolos prensas e feltros, promovendo a consolidação da estrutura da folha, tornando-a uma folha com teor sólido seco em torno de 50%, e em seguida submetê-la a secagem por calor (HENTINEN, 2010). Além disso, a seção pode proporcionar a melhora da lisura da folha de modo a eliminar as marcas deixadas pela tela formadora, aumentar a densidade e a redução do volume específico, melhorando a operacionalidade na área de secagem (HAMAGUCHI, 2007). A capacidade da prensa em desempenhar estas funções, sem causar danos ao papel, como esmagamento, perda de finos, rearranjo das fibras e enrugamento depende do projeto da prensa, do filtro e das características operacionais (NUNES, 2007).

2.2.3.4 Secagem

A água pode estar presente na folha de papel em diversas formas. Pode estar entre as fibras, sobre a superfície das fibras ou nos vazios entre as malhas das fibras. O aumento do grau de refinação das fibras aumenta a dificuldade de secagem da folha, devido ao fato de que com a refinação rompem-se ligações entre as moléculas de celulose havendo novas ligações destas com as moléculas de água (CAMPOS, 2011).

A seção de secagem consiste de um grande número de cilindros rotativos de ferro fundido, aquecidos a vapor, podendo ser seguidos por uma tela sintética permeável denominada tela secadora. Os cilindros são divididos em grupos secadores, onde a pressão do vapor pode ser controlada individualmente, em cada grupo; ao ajustar a pressão do vapor estamos ajustando o fluxo de calor para o papel e consequentemente controlamos a umidade do papel na saída da secagem (AKESSON; SLATTEKE, 2006). O controle de pressão do vapor é proveniente do sistema de vapor e condensado, cuja finalidade é fornecer o vapor aos cilindros secadores, elevar a eficiência da secagem, possibilitar a curva de aquecimento, garantir a drenagem dos secadores e o teor seco do papel (PESCIOL, 2016).

Figura 2.4 – Cilindros de secagem *VALMET*



Fonte: VALMET, 2023

A folha entra na secção secadora a uma temperatura mais baixa que aquela necessária para a evaporação e a fase inicial da secagem será então para elevar a massa envolvida até uma temperatura de evaporação. Esta tarefa fica normalmente a cargo dos primeiros cilindros (normalmente, de 1 a 4), cuja temperatura na superfície do secador é mantida razoavelmente baixa (em torno de 70 a 80 °C), inclusive para prevenir aderência e destacamento de fibras (CAMPOS, 2011).

A maior evaporação específica está situada no centro da área de secagem, pois a massa de água atingiu a temperatura de mudança de fase e toda a energia fornecida pelo sistema é utilizada para este fim. Tão logo a folha de papel esteja aquecida, a temperatura dos cilindros é aumentada rapidamente, podendo chegar ao redor dos 180 °C para alguns papéis menos sensíveis. Ainda assim, a temperatura da folha não irá além da temperatura de evaporação, pois este efeito a esfria, não permitindo que iguale a temperatura do secador (CAMPOS, 2011).

No restante da área de secagem pouco evaporam, pois, nesta região estamos próximos ao limite do processo, onde para pequenos acréscimos no teor seco necessitamos elevado dispêndio de energia, pois a dificuldade de extrair a água fica ainda maior, uma vez que, o que resta de líquido está bem no centro do aglomerado que compõe a folha (CAMPOS, 2011).

Durante a passagem da folha pelos cilindros secadores, a folha é submetida à superfície aquecida dos cilindros, assim a água é evaporada e o papel atinge valores de umidades da ordem de 3% a 8%. Esta evaporação deve ocorrer de forma que irá produzir uma folha seca de alta qualidade, a uniformidade de evaporação no sentido transversal ao da máquina é um dos parâmetros mais crítico. Qualquer variação na evaporação transversal irá produzir uma variação indesejável no perfil de umidade da folha, e outras

propriedades podem ser afetadas, como, por exemplo, a propriedade de tração da folha (PESCIO, 2016).

Segundo Pescio (2016), máquinas projetadas para alta eficiência de secagem, geralmente são providas de uma cobertura em toda a extensão da secagem, denominadas como coifas, com a finalidade de capturar a água evaporada na seção de secagem e removê-la para fora do edifício através de exaustores. O tratamento adequado do fluxo de ar dentro desta seção é de fundamental importância no funcionamento geral da máquina influenciando na eficiência, capacidade de produção, consumo de energia e perfil de umidade.

A quebra da folha durante a operação na seção de secagem é um desafio considerável, resultando em perda de tempo e produtividade nas máquinas de papel de alta velocidade. Essa questão representa uma das principais preocupações enfrentadas durante a operação de máquinas modernas.

2.2.3.5 Colagem superficial

Chama-se colagem superficial a aplicação de produtos químicos tais como amido, colas de dextrina, parafina, álcool polivinílico etc., sobre a superfície do papel já formado e parcialmente seco. Após esta aplicação, com a prensagem (“size press”) da folha a temperaturas adequadas, obtém-se um revestimento que proporciona uma redução nos diâmetros dos poros (espaço entre as fibras) preenchendo estes espaços na folha, dificultando assim a penetração de líquidos, gases ou vapores por esses poros. A colagem superficial é aplicada à folha de papel ou papelão quando ela possui umidade entre 4 e 12%, motivo pelo qual o equipamento é colocado entre grupos de secadores (CAMPOS, 2011).

O “Speed Sizer” (Voith) é um equipamento de colagem que tem conquistado uma firme posição na fabricação de papel, sobretudo quando se deseja, além das características de resistência, características também de superfície e efetuar o tratamento superficial sob altas velocidades. Observa-se que quanto ao “runnability” e às necessidades de secagem, o “speedsizer” supera o “size press” convencional e ainda amplia de forma realmente decisiva o campo de utilização de prensas de colagem convencionais. Esse equipamento baseia-se em dois sistemas de aplicação especiais que depositam a cola, previamente dosada sobre rolos. Ela é transferida ao papel sem a formação de poça no “nip” de prensagem entre os rolos (CAMPOS, 2011).

Figura 2.5 – Modelo *Speed-Sizer* da *VOITH*



Fonte: BACHMEIER, 2014

2.2.3.6 Calandragem

Trata-se de um equipamento da máquina de papel, que é constituído por um conjunto de rolos metálicos, montados uns sobre os outros num cavalete que os sustenta. Consta também, com um dispositivo que permite controlar a pressão de encosto entre os rolos, de forma independente para o lado do acionamento (LA) e o lado do comando (LC), além do centro, através do dispositivo de abaulamento (CAMPOS, 2011).

A principal função da calandra é melhorar as características superficiais da folha, tais como lisura, brilho e auxiliar no controle do perfil de espessura e densidade do papel. Esses efeitos são obtidos através de pressão e fricção entre os rolos, com grande dureza e superfície polida (MENDES, 2006) (ROBUSTI et al., 2014). Existem diversos tipos de calandra, variando desde à quantidade de rolos, configurações até à sua localização na máquina, cada qual possuindo características específicas, vantagens e limitações, que são definidas de acordo com a configuração da máquina, do material processado e do acabamento que se deseja obter no papel (PESCIO, 2016).

2.2.3.7 Enrolamento

Considerado o último elemento da máquina de papel, a seção de enrolamento está associada ao processo contínuo de fabricação e tem a função de transformar a folha em

unidades finitas e independentes, que permitirão o processamento e a utilização do papel. A enroladeira deve produzir bobinas de papel com o maior diâmetro possível e com densidade de enrolamento uniforme, a fim de facilitar a manipulação e a utilização posterior dessas bobinas (ROBUSTI *et al*, 2014).

2.3 INTRODUÇÃO A *SOFT SENSORS*

Nos processos industriais, é comum encontrar uma ampla gama de sensores, como sensores de temperatura, fluxo e pressão, que são projetados para supervisionar, monitorar e controlar em tempo real, garantindo a consistência da qualidade do produto.

Algumas variáveis, que podem ser variáveis de qualidade, por exemplo, não podem ser medidas automaticamente *on-line* devido à falta de sensores ou ao alto custo do sensor, levando à falta de informações suficientes sobre o estado do sistema em tempo real. Normalmente, os testes laboratoriais de amostras de produtos são realizados *offline* para medir a qualidade do produto em uma base de intervalo especificada. Para medir as variáveis de qualidade em tempo real, pode-se usar metodologias de inteligência computacional para construir sensores inteligentes/computacionais para inferir o valor ou as variáveis alvo de qualidade de outras variáveis de processo mensuradas *on-line*. A base para a construção desses sensores inteligentes é que os valores da variável alvo, ou a qualidade do produto, têm uma relação funcional com outras variáveis de processo que podem ser medidas *online*. Esse tipo de sensor inteligente é uma das aplicações de *soft sensors* (FORTUNA *et al*, 2006) (KADLEC *et al*, 2009), aqui referidos como *soft sensors* para aplicações de regressão (SSR) (SOUZA; MENDES, 2016).

Eles são ferramentas importantes para muitos processos industriais, como fábricas de celulose e papel, sistemas de tratamento de águas residuais, fornos de cimento, refinarias e processos de polimerização, apenas para dar alguns exemplos. Em termos gerais, os soft sensors podem ser definidos como modelos inferenciais que usam medições de sensores disponíveis on-line (variáveis fáceis de medir) para estimação on-line de variáveis de qualidade (variáveis difíceis de medir), que não podem ser medidas automaticamente, ou só pode ser medido a um custo elevado, esporadicamente, ou com atrasos elevados (p. ex., análises laboratoriais) (SOUZA; MENDES, 2016).

Em um nível muito geral, pode-se distinguir dois tipos de *soft sensors*, são eles *model-driven* e *data-driven soft sensors*. *Model-driven soft sensors* também são chamados de modelos *white-box* porque eles possuem conhecimento fenomenológico

sobre o histórico do processo. Em contraste com isso, *data-driven soft sensors* são chamados de técnicas *black-box*, porque o modelo em si não tem conhecimento sobre o processo e é baseado em observações empíricas (KADLEC *et al*, 2009).

Modelos *model-driven* (MDM), são desenvolvidos principalmente para fins de planejamento e desenvolvimento das plantas de processo. Estes modelos são baseados em equações que descrevem os princípios químicos e físicos subjacentes ao processo. Um exemplo típico é usar princípios de preservação de massa, equação exotérmica, balanços de energia, cinética de reação na forma de equações de taxa de reação para este fim. A desvantagem deste tipo de modelos é que o seu desenvolvimento requer muito conhecimento especializado do processo e este conhecimento nem sempre está disponível. Outro problema é que os modelos muitas vezes descrevem uma base teórica simplificada do processo, em vez das condições da vida real do processo, que é influenciada por muitos fatores fora do escopo do MDM. Além disso, os modelos orientados por modelo geralmente se concentram na descrição do estado estacionário ideal do processo e, portanto, não são adequados para a descrição de quaisquer estados transitórios (KADLEC *et al*, 2009).

O foco desta revisão é, portanto, colocado nos modelos *data-driven* (DDMs), que surgiram como abordagens de modelagem muito atraentes, aprimorando a caixa de ferramentas de diagnóstico, métodos de prognóstico e de apoio à decisão disponíveis para operadores de instalações e incorporados em sistemas de controlo automatizados. Estes modelos são baseados nas medições reais que são registradas, armazenadas e fornecidas como dados históricos. Os modelos em si são métodos preditivos empíricos como *Principle Component Regression* (PCR), *Multi-layer Perceptron* (MLP), etc. (KADLEC *et al*, 2009).

O desenvolvimento de SSR engloba o mesmo ciclo de projeto dos sistemas de regressão clássicos, no entanto, tem suas próprias peculiaridades. O desenvolvimento do SSR tem as seguintes etapas principais (FORTUNA *et al*, 2006) (KADLEC *et al*, 2009): (a) coleta e filtragem de dados, (b) seleção de variáveis de entrada, (c) escolha e treinamento do modelo, (d) validação do modelo e (e) manutenção do modelo.

Na primeira etapa, os dados são coletados e os objetivos dessa etapa incluem o tratamento de dados ausentes e *outliers*. Os objetivos da segunda etapa são a seleção das entradas mais relevantes, e possivelmente os respectivos desfasamentos de tempo. A escolha e o treinamento do modelo exigem a seleção e o aprendizado corretos do modelo. A etapa de validação do modelo é necessária para julgar se o modelo aprendido reproduz

as variáveis-alvo dentro de níveis aceitáveis de qualidade ou desempenho. O último passo é a manutenção SSR, onde o objetivo é manter uma boa resposta SSR sob a presença de variações de processo ou alguma mudança de dados (SOUZA; MENDES, 2016).

2.3.1 Coleta e tratamentos de dados

As indústrias geralmente precisam armazenar seus dados dos processos. Esta é a base para o uso subsequente de tais dados para otimização do sistema ou outros métodos de transmissão de dados relacionados. Infelizmente, a coleta de dados em aplicações industriais reais vem com problemas bem conhecidos para lidar, como problemas com tempo de amostragem, dados ausentes, *outliers*, condições de trabalho, precisão etc. (SOUZA; MENDES, 2016).

2.3.1.1 Tempo de amostragem

Em sistemas industriais, algumas variáveis são adquiridas em diferentes taxas de tempo. Isso é mais evidente ao analisar as taxas amostrais de variáveis de fácil medição e difícil medição. Na maioria dos problemas, a frequência de aquisição de variáveis fáceis de medir é muito maior, e nesses casos, há a necessidade de sincronizar as variáveis. Este problema é geralmente referido na literatura *como multirate character, ou multiplerate phenomenon* (WU; LUO, 2010). Na prática, as duas abordagens seguintes são as mais comumente adotadas (SOUZA; MENDES, 2016):

1. Redução da amostragem dos dados fáceis de medir, de acordo com a taxa de amostragem mais lenta das variáveis difíceis de medir, excluindo as amostras das variáveis fáceis de medir que não possuem um valor correspondente das variáveis difíceis de medir (alvo) (KADLEC; GABRYS, 2011) (LU *et al*, 2004).
2. Em vez de excluir as amostras que não possuem o respectivo alvo, um modelo de resposta finita ao impulso (FIR) é estimado e aplicado nas amostras para estimar as variáveis difíceis de medir, de baixa taxa de amostragem. A grande preocupação nessa abordagem é a seleção dos valores de ponderação e o comprimento do filtro FIR.

Embora a redução da amostragem por exclusão seja fácil de implementar na prática, ela possui uma desvantagem crítica de perda de informações e pode levar a modelos imprecisos, principalmente se a variável de difícil medição for amostrada de

forma escassa e/ou com atrasos incertos (XIE *et al*, 2013). Uma abordagem melhor é modelar os dados usando o filtro FIR. No entanto, os pesos e o comprimento do filtro FIR devem ser projetados ou estimados com cuidado (SOUZA; MENDES, 2016).

2.3.1.2 Dados ausentes

É bastante comum ter observações com valores ausentes para uma ou mais variáveis. O problema dos dados ausentes ocorre quando nenhum valor é armazenado para uma variável em uma observação. Existem duas abordagens comuns para lidar com dados ausentes. A primeira é a remoção das amostras que contêm dados ausentes, uma abordagem também conhecida como exclusão completa de casos. A segunda abordagem é preencher os valores ausentes usando algum método de imputação. A primeira abordagem pode ser utilizada se o número de valores ausentes for pequeno, mas caso contrário, deve ser evitada (HASTIE *et al*, 2001). No segundo caso, a estratégia mais simples é imputar o valor ausente com a média ou mediana dos valores não ausentes para aquela variável. Outra abordagem é a imputação por *hot-deck*, onde um valor ausente é imputado a partir de um valor selecionado aleatoriamente dos valores de entrada para valores-alvo semelhantes (ANDRIDGE; LITTLE, 2010). Esses métodos de imputação por média/mediana e imputação por *hot-deck* são geralmente referidos como múltipla imputação (SOUZA; MENDES, 2016).

Dois outros métodos frequentemente utilizados para lidar com dados ausentes são o método da máxima verossimilhança (ML) e o método da maximização da expectativa (EM). O método ML modela a(s) variável(is) ausente(s) com base nos dados disponíveis. Essencialmente, o ML assume algum modelo para a distribuição dos dados da variável ausente e, em seguida, os parâmetros do modelo são estimados (SOUZA; MENDES, 2016). O método EM, por sua vez, é uma abordagem iterativa que estima tanto os valores ausentes quanto os parâmetros do modelo. Ele parte de uma estimativa inicial dos valores ausentes e dos parâmetros do modelo e, em seguida, itera entre as etapas de estimativa dos valores ausentes e dos parâmetros até que a convergência seja alcançada.

2.3.1.3 *Outliers*

Outliers são valores de observação que se desviam significativamente do intervalo típico e significativo de valores. As observações assumem valores inconsistentes quando

comparadas à maioria dos dados registrados, e isso pode afetar grandemente o desempenho do projeto do SSR (KADLEC *et al*, 2009).

Os *outliers* podem ser causados, por exemplo, por mau funcionamento do sensor, erros de comunicação ou degradação do sensor. Para mitigar os efeitos dos *outliers*, é necessário primeiro detectá-los e, em seguida, tratá-los. No entanto, ao aplicar métodos de detecção, os resultados geralmente precisam ser validados manualmente pelo desenvolvedor do modelo e/ou pelo especialista do processo. O objetivo da inspeção manual é detectar possíveis mascaramentos de *outliers* (ou seja, detecções falsas negativas – *outliers* não detectados) e o fenômeno de inundação de *outliers* (ou seja, detecções falsas positivas – valores corretos rotulados como *outliers*) (KADLEC *et al*, 2009).

Os métodos mais comuns de detecção de *outliers* são baseados em técnicas estatísticas. A abordagem mais simples é a regra 3σ (PEARSON, 2002), que se baseia em uma distribuição univariada das variáveis. A regra 3σ funciona da seguinte forma: assumindo que uma variável é extraída de uma distribuição gaussiana com média μ e desvio padrão σ , as amostras dessa variável que estão fora dos limites $[\mu-3\sigma, \mu+3\sigma]$ são consideradas *outliers*. Uma versão robusta da regra 3σ é o identificador *Hampfel* (DAVIES; GATHER, 1993), que leva em consideração a média absoluta e o desvio absoluto médio. O identificador *Hampfel* é adequado em casos em que os dados são severamente afetados por *outliers*, e tem se mostrado efetivo na prática em aplicações reais (LIU *et al*, 2004) (PEARSON, 2002).

2.3.2 Seleção de variáveis de entrada

Nas aplicações de SSR, frequentemente há uma grande quantidade de candidatos a variáveis de entrada provenientes da estrutura de supervisão do processo. O número de candidatos pode variar para milhares.

O uso de modelos *black-box* já sugere que o projetista de SSR possui pouco conhecimento sobre o sistema a ser modelado e, consequentemente, sobre as variáveis que afetam a variável alvo. No entanto, isso não é verdade em todos os casos, uma vez que, na maioria das aplicações de SSR, a seleção de um conjunto das variáveis mais relevantes é feita por especialistas do sistema. No entanto, para processos fisicamente grandes e altamente integrados, a enumeração e seleção de variáveis candidatas com base no conhecimento do processo pode não ser viável (WARNE *et al*, 2004).

Além disso, a maioria dos trabalhos na literatura indica que frequentemente apenas algumas variáveis são necessárias para compor o modelo SSR. Um número reduzido de variáveis tem várias vantagens, como a redução do tempo de desenvolvimento do modelo, a possibilidade de agregação das informações sobre a interpretação física do processo ou a melhoria do desempenho do modelo. Além disso, a redução do número de variáveis implica em um menor número de sensores reais necessários, reduzindo custos e aumentando ou viabilizando aplicações (SOUZA; MENDES, 2016).

Stepwise regression é amplamente utilizada para seleção de variáveis em regressão linear. O procedimento constrói iterativamente uma sequência de modelos de regressão, adicionando ou removendo variáveis a cada passo. O critério para adicionar ou remover uma variável em cada etapa é geralmente expresso em termos de um teste *F* parcial. Seja F_{in} o valor da variável aleatória *F* para adicionar uma variável ao modelo e seja F_{out} o valor da variável aleatória *F* para remover uma variável do modelo. É necessário que F_{in} seja maior do que F_{out} para tornar relativamente mais difícil adicionar uma variável do que remover uma. O procedimento padrão de *stepwise regression* pode ser resumido da seguinte forma (MONTGOMERY *et al*, 2001):

A regressão começa formando um modelo de uma variável usando a variável regressora que possui a maior correlação com a variável de resposta *Y*. Essa também será a variável regressora que produz o maior valor de estatística *F*. Por exemplo, suponha que nesta etapa x_1 seja selecionada. Na segunda etapa, as $(K - 1)$ variáveis candidatas restantes são examinadas, e a variável para a qual a estatística *F* parcial

$$F_j = \frac{SS_R(\beta_j | \beta_1, \beta_0)}{MS_E(x_j, x_1)} \quad (2.1)$$

é máxima é adicionada à equação, desde que $f_j > f_{in}$. Na Equação 2.1, $MS_E(x_j, x_1)$ representa a média dos quadrados para o erro do modelo que contém tanto x_1 quanto x_j . Suponha que esse procedimento indique que x_2 deve ser adicionado ao modelo. Agora, o algoritmo determina se a variável x_1 adicionada na primeira etapa deve ser removida. Isso é feito calculando a estatística *F*:

$$F_1 = \frac{SS_R(\beta_1 | \beta_2, \beta_0)}{MS_E(x_1, x_2)} \quad (2.2)$$

Se o valor calculado $f_1 < f_{out}$ a variável x_1 é removida; caso contrário, ela é mantida, e tentaríamos adicionar um regressor ao modelo contendo tanto x_1 quanto x_2 .

Em geral, em cada etapa, o conjunto de regressores candidatos restantes é examinado, e o regressor com a maior estatística F parcial é incluído, desde que o valor observado de f seja superior a f_{in} . Em seguida, a estatística parcial F é calculada para cada regressor no modelo, e o regressor com o menor valor observado de F é removido se o valor observado $f < f_{out}$. O procedimento continua até que nenhum outro regressor possa ser adicionado ou removido do modelo (MONTGOMERY *et al*, 2001).

Stepwise regression é quase sempre realizada usando um *software* matemático.

2.3.3 Escolha e treinamento do modelo

Para sistemas complexos, a abordagem de modelagem *white-box* pode ser praticamente inviável. Os modelos *black-box* ou *data-driven* são baseados em observações empíricas do processo (os próprios métodos são métodos empíricos de previsão).

A modelagem *black-box* é capaz de descrever as condições reais do processo e requer pouco conhecimento sobre o sistema a ser modelado. No entanto, exige um trabalho intensivo com os dados do processo. Algumas dificuldades desses tipos de abordagens estão relacionadas à dificuldade de escolher o tipo e a estrutura corretos do modelo, as funções a serem usadas e a quantidade de termos de função necessários para o desenvolvimento (SOUZA; MENDES, 2016).

Na modelagem *black-box*, o primeiro aspecto a ser decidido é qual tipo de modelo será utilizado. Existem sempre duas opções: um modelo linear ou um modelo não linear. De acordo com muitos autores, um modelo linear deve ser sempre considerado antes de um modelo não linear. Se o modelo linear não fornecer resultados satisfatórios, uma possível explicação, entre muitas outras possibilidades, é que o sistema possui um comportamento não linear, sendo assim um modelo não linear deve ser a melhor escolha (NELLES, 2001).

2.3.4 Validação do modelo

O objetivo da etapa de validação do modelo é avaliar a capacidade/habilidade do modelo treinado de realizar generalização para novas amostras. A precisão da

generalização também pode ser usada como um estimador para a classificação do modelo em uma abordagem de seleção de variáveis (por exemplo, na seleção de variáveis envolvida) (SOUZA; MENDES, 2016).

Para um grande conjunto de dados, o modelo geralmente é aprendido usando apenas uma parte dos dados e, em seguida, o desempenho do modelo é medido nos dados restantes, geralmente chamados de conjunto de dados de validação, usando alguma métrica de desempenho, geralmente o MSE (por exemplo, valores menores de MSE indicam melhores modelos) ou o erro quadrático médio normalizado (NRMSE). O NRMSE é uma versão normalizada do MSE, frequentemente expressa em porcentagem, que fornece uma análise mais intuitiva do desempenho do modelo (SOUZA; MENDES, 2016).

Para conjuntos de dados pequenos, uma técnica de validação cruzada geralmente é empregada para avaliar o desempenho do modelo. As técnicas comuns de validação cruzada são a validação cruzada com *K-fold* e a validação cruzada *leave-one-out* (LOOCV) (SOUZA; MENDES, 2016).

2.3.5 Manutenção do *soft sensors*

Durante o projeto do SSR, os dados históricos do processo são usados para aprender o modelo. No entanto, os dados históricos contêm informações limitadas, correspondendo a um período de tempo limitado e possivelmente focando em um conjunto limitado de áreas de operação do espaço de estados (SOUZA; MENDES, 2016).

Ao lidar com eventos novos, não descritos nos dados históricos, o SSR tende a diminuir seu desempenho. Nesse contexto e para superar essa deterioração de desempenho, o objetivo da manutenção do SSR é manter uma boa resposta do SSR mesmo na presença de variações do processo ou alguma mudança nos dados. Geralmente, isso é feito atualizando o modelo SSR online/recursivamente, em modo em lote ou amostra, usando as amostras recebidas do processo (neste contexto, os SSRs são chamados de "SSRs adaptativos") (KADLEC *et al*, 2011).

Do ponto de vista do aprendizado de máquina, a área dos SSRs adaptativos está relacionada ao problema da mudança de conceito. Mudança de conceito significa que as propriedades estatísticas da variável alvo mudam ao longo do tempo, sendo que o termo conceito se refere ao objeto/alvo a ser previsto (SOUZA; MENDES, 2016). Existem três abordagens comumente empregadas para lidar com a mudança de conceito: (1) seleção

de amostras, (2) ponderação de amostras e (3) aprendizado em conjunto (ou aprendizado com múltiplos descritores de conceito) (TSYMBAL, 2004).

Além disso, como discutido anteriormente, os modelos mais utilizados em aplicações de *soft sensors* são baseados em métodos estatísticos multivariados (*stepwise*, PCA) ou técnicas de inteligência artificial (redes neurais). Em *soft sensors* adaptativos, tais modelos também podem ser utilizados, mas existem preocupações em relação ao aprendizado/adaptação de parâmetros. Os modelos podem ser aplicados como um único modelo, nas abordagens de ponderação ou seleção de amostras, ou vários modelos podem ser aplicados juntos na abordagem de conjunto (SOUZA; MENDES, 2016).

2.4 REDES NEURAIS ARTIFICIAIS

De modo geral, as Redes Neurais Artificiais (RNA) são estruturas paralelas, maciçamente distribuídas, constituídas por unidades simples de processamento conhecidas como neurônios (HAYKIN, 2001). A partir dessas unidades, as RNAs possuem a capacidade de aprender de acordo com o problema no qual é aplicada, através de ajustes nas conexões que são chamadas de pesos sinápticos, que ficam entre as camadas de entrada e saída da rede (RODRIGUES, 2022).

As redes inicialmente recebem dados de entrada, que forma a primeira camada, e então os dados são aplicados a uma determinada função, normalmente não linear, para se obter dados de saída (última camada). Todas as camadas entre a primeira e a última são chamadas de camadas ocultas. Os pesos, que fazem parte da camada oculta, são responsáveis por armazenar o conhecimento adquirido pelo modelo, e por ponderar a entrada recebida por cada neurônio, de modo a determinar se esta fará parte do aprendizado ou não (RODRIGUES, 2022).

A capacidade de aprender através de exemplos, ter saídas coerentes quando apresentada a dados nunca visto antes, e de generalizar este aprendizado é o principal atrativo do uso de RNAs. Esta generalização é uma demonstração do que as mesmas podem fazer, e vai muito além do que simplesmente mapear relações entre dados de entrada e saída. As RNAs são capazes de extrair informações significativas em uma grande quantidade de dados, e no caso de problemas industriais, buscar instabilidades no processo com o mínimo de tempo e custo e máxima precisão (MORAIS, 2010).

Para se obter o aprendizado desejado de um problema a partir do uso de RNAs, uma série de cálculos iterativos devem ser feitos de forma a ajustar os parâmetros ideais,

que são os pesos e *bias*, para que no final do processo o conhecimento que a rede adquiriu seja guardado e posteriormente utilizado em dados novos do processo, nunca vistos pela rede. Esses procedimentos caracterizam os algoritmos de aprendizado (BRAGA, CARVALHO & LUDELMIR, 2000).

2.4.1 Tipos de redes neurais artificiais

Em geral, existem três classes fundamentais de arquiteturas de RNAs: rede *feedforward* de única camada, rede *feedforward* de múltiplas camadas e redes recorrentes. As duas primeiras se diferem pelo fato de uma ter apenas entrada e saída, sem presença de camada intermediária, enquanto a outra possui camadas intermediárias, que a torna mais complexa de acordo com a quantidade de camadas ocultas inseridas. Ambas possuem o fluxo de informação apenas em uma direção, que seria desde a entrada até a saída. Como a informação se move de forma unidirecional através da rede, ela nunca passa duas vezes no mesmo neurônio (ARAÚJO NETO, 2018).

As redes *feedforward* são bastante utilizadas, porém, se tratando de resolver problemas com dados sequenciais, as mesmas não apresentam um bom funcionamento, visto que não possuem armazenamento de conhecimento adquirido anteriormente por uma entrada, para lidar com outra que seja sequencial (RODRIGUES, 2022).

Redes neurais recorrentes (RNNs) são sistemas dinâmicos; elas possuem um estado interno em cada etapa de classificação. Isso ocorre devido às conexões circulares entre neurônios de camadas superiores e inferiores, e conexões de retroalimentação opcional. Essas conexões de retroalimentação permitem que as RNNs propaguem dados de eventos anteriores para etapas de processamento atuais. Assim, as RNNs constroem uma memória de eventos de séries temporais (STAUDEMEYER; MORRIS, 2019).

As RNNs trazem as informações dadas em uma entrada da rede, para obter uma saída, e quando a rede ler a próxima entrada ela carrega com si o conhecimento adquirido na etapa anterior, e assim por diante até o final dos dados sequenciais. De forma esquemática, temos que a rede ler uma *feature* de um input alimentado a uma rede neural e gera um *output*. Quando a rede lê a próxima *feature*, traz consigo a informação do passo anterior para poder gerar o próximo *output* e assim por diante (GOMAA, 2020).

As redes recorrentes são diferentes das redes *feedforward* pelo *loop* de *feedback* conectado às suas saídas anteriores, ingerindo as mesmas passo a passo como entrada da seguinte. Essas informações sequenciais são armazenadas nas camadas ocultas da RNN,

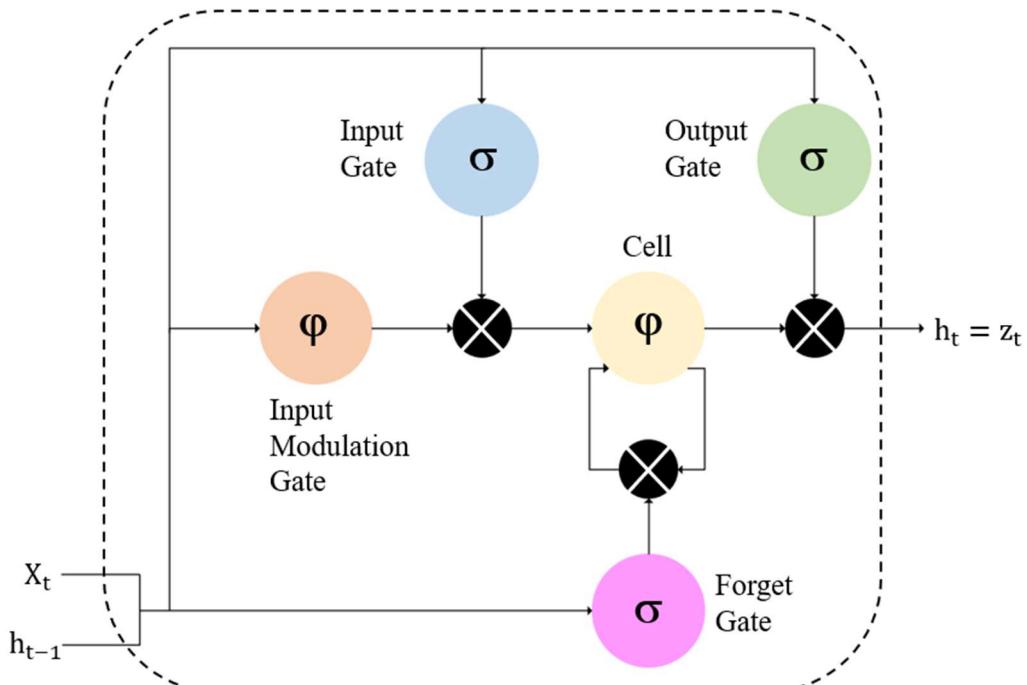
que consegue passar por muitas etapas de tempo à medida que ela avança em cascata para afetar o processamento de cada novo exemplo (RODRIGUES, 2022).

2.4.2 Long Short-Term Memory

A *LSTM* é um tipo de rede neural capaz de armazenar valores em intervalos arbitrários, sendo adequada para classificar, processar e prever séries temporais com intervalos de tempos desconhecidos. Desse modo, utiliza-se a *LSTMs* quando a rede “esquece” o que aprendeu anteriormente e não consegue propagar as informações através da sequência inteira. Essa abordagem é a razão dessa rede ganhar espaço em plantas industriais pelo fato de conectar informações anteriores à atual e fazer com que essas informações persistam durante o treinamento (RODRIGUES, 2022).

As RNN do tipo *LSTM* possuem uma estrutura diferente das demais RNN. Nesse tipo, a estrutura é formada por um módulo de repetição, onde existem quatro camadas e uma célula que compõe o bloco de memória da rede. As informações aprendidas pela rede são retidas pelas células e as manipulações de memória são feitas pelos portões, chamados de *gates* (CASTELÃO, 2018). A Figura 2.6 mostra o esquema de uma rede *LSTM*.

Figura 2.6 – Esquema representativo de uma rede recorrente do tipo *LSTM*



Fonte: Adaptado de GOMAA, 2020

Como pode ser visto na Figura 2.6, e como descreve Rodrigues (2022), as *LSTM* possuem os *gates*, que são responsáveis por proteger e controlar o estado da célula, de modo a permitir ou bloquear que a informação passe a ser aprendida pela rede.

A primeira porta é a *Forget Gate*, que remove as informações que não são mais úteis no estado da célula. Nela, duas entradas são inseridas, sendo X_t a informação nova e h_{t-1} a saída da célula anterior. Nessa porta, as entradas são multiplicadas por matrizes de zeros e somadas ao *bias* (valor constante que ajuda a controlar o comportamento das unidades), passam por uma função de ativação (função que permite o aprendizado e representa as relações complexas nos dados) e tem como resposta uma saída binária, onde 0 corresponde a resposta ser esquecida e 1 a ser retida para uso futuro.

A segunda porta, chamada de *Input Gate* é onde as informações são adicionadas ao estado da célula. Primeiro, a informação passa por uma função sigmoide que filtra os valores a serem lembrados de forma similar ao *Forget gate* usando as entradas X_t e h_{t-1} . Então, um vetor é criado usando a função $\tan(h)$ que tem como saída o intervalo de -1 a +1, que contém todos os valores possíveis de X_t e h_{t-1} . Os valores do vetor e os valores regulados são multiplicados para obter as informações úteis.

O *Output gate* tem como objetivo extrair as informações úteis para o estado da célula que serão apresentadas como saída da rede. Primeiro, um vetor é gerado aplicando a função $\tan(h)$ na célula. Então, a informação é regulada usando a função sigmóide que filtra os valores a serem lembrados usando as entradas X_t e h_{t-1} . Os valores do vetor e os valores regulados são multiplicados para serem enviados como uma saída e entrada para a próxima célula.

As setas em *loop* indicam a natureza recursiva da célula. Isso permite que as informações sejam armazenadas na célula *LSTM*. É essa estrutura que permite que a mesma seja utilizada em problemas sequencias com dependências de longo prazo, como séries temporais semelhantes ao problema deste trabalho.

3 METODOLOGIA

3.1 HISTÓRICO E COLETA DE DADOS

A criação de uma rede neural requer um histórico de dados adequado e uma coleta precisa para garantir o desempenho e a eficácia do modelo. O histórico de dados é essencial, pois fornece informações sobre o comportamento do sistema ao longo do tempo

e permite que a rede neural aprenda com padrões passados. Para construir uma rede neural eficaz, é importante coletar dados relevantes e representativos do problema em questão.

Os dados utilizados para criar a rede neural foram extraídos do PI *Processbook* da Suzano, unidade Industrial Mucuri. O PI *Processbook* foi utilizado como ferramenta para coletar e armazenar dados históricos das operações da Suzano, permitindo análises mais detalhadas e a criação de um modelo preditivo por meio da rede neural. Esses dados históricos da Suzano foram essenciais para o desenvolvimento de um sistema de previsão e otimização baseado em inteligência artificial.

A rede neural foi especialmente projetada para prever a gramatura do papel na indústria de papel e celulose, com foco na empresa Suzano S.A. A gramatura do papel é uma variável essencial que determina a densidade e a espessura do papel, afetando diretamente na qualidade das entregas do setor.

Durante o desenvolvimento da rede neural, o principal objetivo foi treiná-la para fazer previsões precisas da gramatura do papel com base em produções passadas. Foram coletados e utilizado um histórico de dados, incluindo informações sobre desempenho da máquina, consistências da massa, fração de carga mineral, vazões de processo e pressão da caixa de vapor. Esses dados serviram como entrada para treinar a rede neural, permitindo que ela aprendesse os padrões e as relações entre as variáveis de entrada e a gramatura do papel.

O desempenho da máquina de produção, incluindo parâmetros como velocidade, pressão de formação e secagem, influencia a compactação das fibras do papel, enquanto que as consistências da massa, tanto na entrada quanto na saída do processo, afetam a formação da folha de papel e têm um impacto direto na gramatura. A fração de carga mineral adicionada à suspensão de celulose também desempenha um papel importante, podendo influenciar a densidade e o peso do papel.

Além disso, as vazões de processo, como a vazão da caixa de entrada e pareto, tem influência na distribuição das fibras e, consequentemente, na gramatura do papel. A pressão da caixa de vapor, relacionada à etapa de secagem, afeta a taxa de evaporação da água presente na folha de papel, o que influencia sua densidade e gramatura.

As relações entre essas variáveis e a gramatura do papel podem ser complexas e não lineares. Por isso, o uso de uma rede neural é benéfico, pois permite capturar essas relações não lineares e realizar previsões mais precisas.

Considerando diferentes aspectos relacionados à qualidade dos dados utilizados, um dos desafios enfrentados foi lidar com *outliers*, que podem surgir devido a diversas

fontes, como quebras de papel durante o processo, paradas programadas e não programadas de máquina, bem como erros na leitura dos sensores.

A presença de *outliers* causados por quebras de folha e paradas de máquina pode afetar significativamente a qualidade dos dados utilizados no treinamento da rede neural. Além disso, os erros na leitura dos sensores também representam um desafio, uma vez que podem introduzir dados incorretos ou superestimados, e podem ter sido causados por falhas técnicas, ruídos ou problemas de calibração dos sensores. Esses eventos anômalos podem introduzir variações abruptas e não representativas na gramatura do papel, o que pode comprometer a precisão das previsões.

A remoção desses dados visou melhorar a qualidade e a confiabilidade das entradas utilizadas no treinamento e na previsão, fortalecendo a capacidade da rede neural em lidar com situações anômalas e garantindo um desempenho mais robusto em diferentes cenários.

3.2 MÉTODO DE SELEÇÃO DE VARIÁVEIS DE ENTRADA

Para a seleção de variáveis foi utilizado o método *stepwise regression*. Esse método é uma abordagem iterativa que envolve a inclusão e exclusão sistemática de variáveis com base em critérios estatísticos. Inicialmente, todas as variáveis disponíveis foram incluídas no modelo. Em seguida, o método *stepwise regression* foi aplicado para selecionar as variáveis mais relevantes, removendo aquelas que apresentavam baixa contribuição para a previsão da gramatura.

Stepwise regression utiliza critérios como o valor-p, a estatística F e medidas de desempenho do modelo para avaliar a importância das variáveis. Durante cada iteração, o método realiza testes estatísticos para determinar se a inclusão ou exclusão de uma variável melhora o desempenho geral da rede neural.

O processo de seleção de variáveis continua até que um critério pré-definido seja atingido, como um determinado valor de significância estatística ou uma medida de desempenho satisfatória. Isso garante que apenas as variáveis mais relevantes e significativas sejam incluídas no modelo final.

Ao aplicar o método, foram considerados não apenas os critérios estatísticos, mas também o conhecimento do domínio e a relevância das variáveis em relação à gramatura do papel. Isso permitiu a construção de um modelo mais preciso e interpretável, ao mesmo tempo em que reduziu a complexidade ao utilizar apenas as variáveis mais importantes.

Para realizar a regressão *stepwise*, foi utilizada a função do *MATLAB*, uma ferramenta amplamente utilizada em análises estatísticas e modelagem de dados.

```
clc
clear all

%Importando a tabela do arquivo Excel
dataset = readtable('C:\Downloads\base-de-dados.xlsx')
```

A base de dados foi importada de um arquivo *Excel*, e as colunas relevantes da tabela são extraídas e armazenadas em variáveis separadas, como $x_1, x_2, x_3, \dots, x_{11}$. Cada variável representa uma característica ou atributo que será considerado na análise.

```
%Extraindo dados das colunas da tabela da base de dados
x1 = table2array(dataset(:, 'PRESS-DEVRONIZER'))
x2 = table2array(dataset(:, 'VELOCIDADE'))
x3 = table2array(dataset(:, 'TEOR-CINZAS'))
x4 = table2array(dataset(:, 'CONC-CXE'))
x5 = table2array(dataset(:, 'CONC-AGUA-BRANCA'))
x6 = table2array(dataset(:, 'TAXA-PRODUCAO'))
x7 = table2array(dataset(:, 'CONC-REFINACAO'))
x8 = table2array(dataset(:, 'CONC-TQ-MISTURA'))
x9 = table2array(dataset(:, 'CONC-CX-MAQUINA'))
x10 = table2array(dataset(:, 'VAZAO-PARETO'))
x11 = table2array(dataset(:, 'VAZAO-CXE'))
```

Após a extração das colunas, é criada a matriz de design X e o vetor de resposta Y . A matriz X é formada pela concatenação das variáveis extraídas, juntamente com os valores de x_0 que representa o termo de interceptação. O vetor Y contém os valores da coluna gramatura da tabela, que representa a variável dependente.

```
%Aplicando Stepwise Regression
x0= ones(size(x1))

X= [x0,x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11]
Y= table2array(dataset(:, 'GRAMATURA'))

swgram= stepwisefit(X,Y, 'PEnter', 0.2)
```

Por fim, a função *stepwisefit* é utilizada para realizar a regressão. Essa função recebe a matriz de design X , o vetor de resposta Y e um parâmetro chamado '*PEnter*', que

define o limite de significância para inclusão de variáveis no modelo. O resultado da regressão é armazenado na variável *swgram*, que conterá informações sobre quais variáveis foram selecionadas como relevantes para a rede neural.

Uma das principais vantagens do método é a simplificação do modelo, removendo variáveis desnecessárias e evitando problemas de multicolinearidade. É importante considerar atentamente os critérios estabelecidos e interpretar os resultados com cautela.

3.3 ESTRUTURA DO MODELO LSTM PROPOSTO

A principal motivação para o desenvolvimento do modelo *Long Short-Term Memory* foi superar as limitações das RNNs convencionais para lidar com o desafio de capturar dependências de longo prazo em sequências temporais.

LSTM pode aprender como superar atrasos mínimos de tempo de mais de 1.000 etapas de tempo discretas. O acesso às células é controlado por unidades de portão multiplicativo (como o portão de esquecimento e o portão de entrada), que aprendem quando conceder acesso (STAUDEMAYER; MORRIS, 2016), permitindo ao modelo controlar o fluxo de informações.

Primeiramente, é feito o carregamento da base de dados a partir de um arquivo *excel*, que contém informações de tempo e dados relacionados com a gramatura. Para isso, utiliza-se a biblioteca do *pandas* do *python*.

```
#importando biblioteca pandas
import pandas as pd

#importando base de dados
dataset = pd.read_excel('C:\Downloads\base-de-dados.xlsx')
gramatura = dataset['GRAMATURA'].to_numpy().reshape(-1,1)
```

A próxima etapa consiste na preparação dos dados de entrada e na divisão dos dados em conjuntos de treinamento e teste. O conjunto de treinamento é obtido a partir de 70% dos dados, enquanto o conjunto de teste é formado pelos 30% restantes.

Essa divisão é realizada de forma estratégica, a fim de garantir uma avaliação objetiva do modelo. O conjunto de treinamento é utilizado para ajustar os parâmetros do *soft sensor* e para realizar o processo de aprendizado, permitindo que o modelo se adapte aos padrões e relações presentes nos dados. Por sua vez, o conjunto de teste é utilizado

para avaliar o desempenho do modelo em dados não vistos anteriormente, verificando sua capacidade de generalização e previsão em situações reais.

```
#criando matriz de dados de entradas
inputs = dataset.drop('GRAMATURA',axis=1)

#dividindo os dados em treinamento e teste
data_length = int(len(gramatura)*0.7)
data_train, data_test = gramatura[:data_length], gramatura[data_length:]
```

Após a divisão dos dados, é realizada a normalização dos valores da gramatura usando a classe *MinMaxScaler* do *sklearn*. Essa etapa é importante para garantir que os dados estejam na mesma escala e facilitem o treinamento do modelo. Em seguida, os dados de treinamento e teste são preparados para alimentar a rede.

```
#importando as bibliotecas numpy e sklearn
import numpy as np
from sklearn.preprocessing import MinMaxScaler

#normalizando os dados
scaler = MinMaxScaler(feature_range=(0,1))
scaler_train = scaler.fit_transform(data_train)
scaler_test = scaler.transform(data_test)

scaler_inputs = MinMaxScaler(feature_range=(0,1))
scaler_inputs.fit(inputs)
scaled_inputs = scaler_inputs.transform(inputs)
```

Os dados de treinamento são organizados em sequências e separados em entradas e saídas esperadas. O mesmo é feito para os dados de teste.

```
#importando a biblioteca keras
from keras.models import Sequential
from keras.layers import Dense, LSTM, Dropout

#construindo o modelo LSTM --- rede neural
model = Sequential()
model.add(LSTM(50, return_sequences=True, input_shape=(
    1,x_train.shape[2])))
model.add(LSTM(50, return_sequences=False))
model.add(Dense(25))
model.add(Dense(1))
```

O modelo LSTM é construído utilizando a biblioteca *keras*. É definida uma sequência de camadas, com duas camadas LSTM com 50 neurônios cada, seguidas por uma camada densa de 25 neurônios e camada de saída. O modelo é compilado com o otimizador “*adam*” e a métrica de perda utilizada é o erro quadrático médio para a construção e treinamento do modelo de aprendizado de máquina.

Em seguida, o modelo é treinado usando os dados de treinamento. O treinamento é realizado em lotes (*batch_size*) e ao longo de um número de épocas (*epochs*) definido. Durante o treinamento, também é calculada a perda nos dados de validação.

```
#compilando o modelo para treinamento
model.compile(optimizer="adam", loss="mse", metrics=['mae'])

#treinando o modelo com os dados de treinamento (x_train,y_train)
results = model.fit(x_train,y_train,batch_size=100,epochs=30,
                     validation_split=0.3)

#fazendo previsões com os dados de teste (x_test)
model_predict = model.predict(x_test)
model_predict = scaler.inverse_transform(model_predict)
```

Após o treinamento, o modelo é utilizado para fazer previsões nos dados de teste. As previsões são desnormalizadas usando o *scaler* inverso para obter os valores de gramatura reais.

Finalmente, são criados os dados necessários para plotar os resultados. Os dados de treinamento e as previsões do modelo são organizados em um *dataframe*, que será usado para visualizar a série temporal original e as previsões do modelo em um gráfico.

4 RESULTADOS E DISCUSSÕES

Para a reprodução dos resultados a seguir, foi utilizado um conjunto de dados com mais de 190 mil registros, coletados ao longo dos meses de março e abril de 2023.

Antes de prosseguir com a criação do *soft sensor*, foi realizada uma etapa de limpeza dos dados, a fim de garantir a qualidade e a consistência das informações. Durante essa etapa, foram identificados e tratados eventuais dados faltantes, *outliers* e inconsistências, de modo a evitar interferências indesejadas nos resultados finais. Para esse fim, uma abordagem estatística baseada na análise de significância das variáveis disponíveis no conjunto de dados foi realizada utilizando *stepwise regression*. Com base

nos resultados obtidos, as variáveis que possuem uma influência significativa no processo foram identificadas e utilizadas como entradas para o modelo *LSTM*.

Tabela 4.1 – Resultados da regressão *stepwise* feita no *Matlab*

variável	unidade	coeff.	standard error	status	P-value
pressão devronizer	bar	-3,4317	0,1248	In	3,1e-162
velocidade da máquina	m/min	-0,1001	0,000199	In	0
teor de cinzas	%	-0,1288	0,0077	In	8,43e-62
concentração cx de entrada	g/L	-0,02	0,0151	In	0,1866
concentração água branca	g/L	0,3101	0,1347	In	0,0214
taxa de produção	t/h	2,2199	0,0064	In	0
concentração refinação	%	0,1508	0,1869	Out	0,4196
concentração tq de mistura	%	-1,6876	0,2238	In	5,01e-14
concentração cx da máquina	%	-2,9714	0,2576	In	1,2e-30
vazão do pareto	L/min	0,0076	0,0015	In	7,03e-07
vazão caixa de entrada	m ³ /min	0,000175	0,0047	Out	0,9701

A Tabela 4.1 apresenta os resultados da análise na qual foram avaliadas diversas variáveis em relação à gramatura.

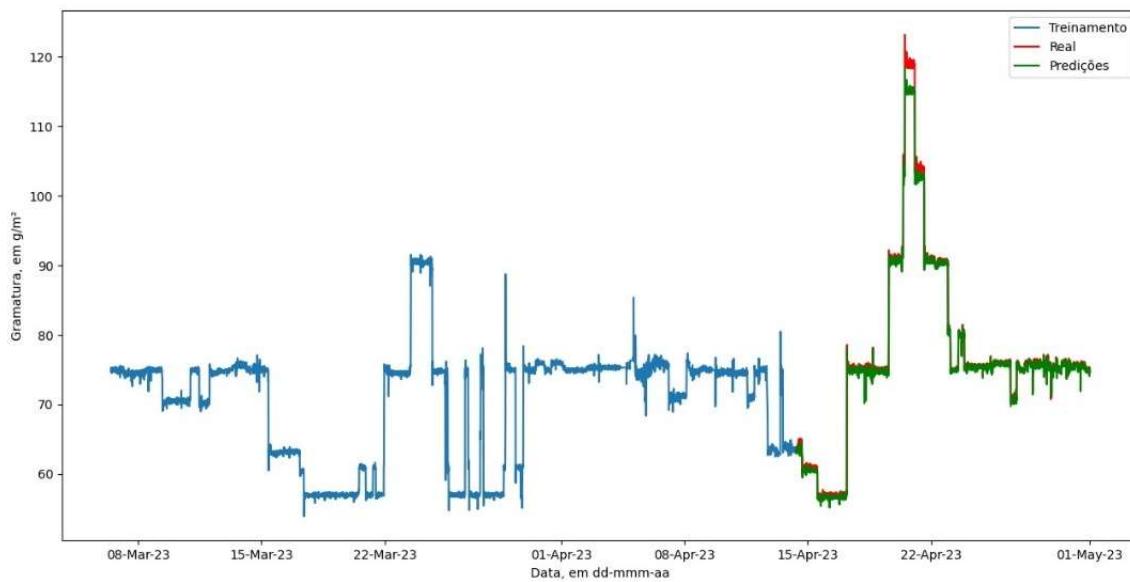
Os coeficientes estimados indicam a magnitude e a direção da relação entre essas variáveis e a gramatura. Por exemplo, um coeficiente negativo para a velocidade da máquina sugere que um aumento nessa variável está associado a uma redução na gramatura. Por outro lado, um coeficiente positivo para a taxa de produção indica que à medida que a gramatura aumenta, valores maiores para a variável são esperados. O erro-padrão é uma medida da precisão da estimativa do coeficiente; quanto menor o erro-padrão, mais precisa é a estimativa; isso significa que a variável com um erro-padrão menor tem uma estimativa do coeficiente mais confiável.

Além disso, os valores de P fornecem uma medida da significância estatística da relação entre a variável e a gramatura. Para variáveis com valores de P baixos, significa que essas entradas são importantes para explicar as variações durante o processo e devem ser consideradas ao otimizar o modelo. As variáveis marcadas como "In" na coluna *status* foram consideradas significativas, enquanto que as variáveis marcadas como "Out" foram desconsideradas da base de dados. Ao descartar essas variáveis do conjunto, podemos simplificar o modelo, sem comprometer a capacidade de previsão.

Aplicando-se as variáveis avaliadas ao treinamento, foi possível melhorar o desempenho e a capacidade de aprendizado do modelo. A inclusão apenas das variáveis

estatisticamente significativas proporcionou um ajuste mais eficiente aos dados, eliminando o ruído e aumentando a relevância das informações utilizadas. Essa estratégia resultou em um modelo mais robusto, capaz de fazer previsões mais confiáveis da gramatura ao considerar apenas as variáveis mais relevantes.

Figura 4.1 – Evolução do treinamento com comparação entre previsões e valores reais



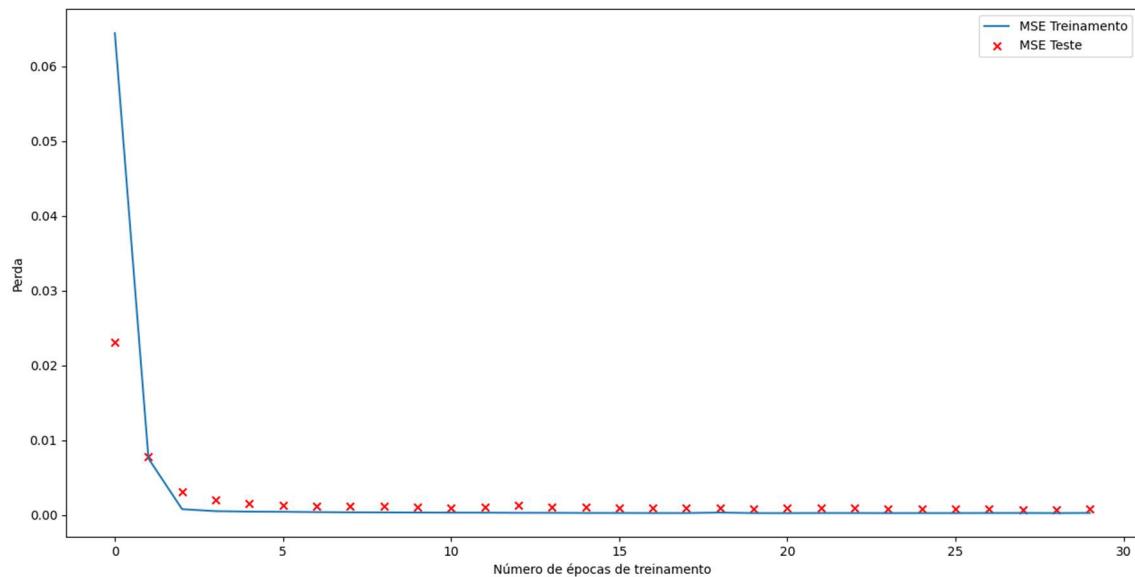
Analizando o gráfico, podemos perceber que o modelo apresenta um bom desempenho na predição de baixas gramaturas. No entanto, é importante ressaltar que o modelo apresenta uma menor precisão durante o treinamento quando os valores ultrapassam a escala de 100 g/m² (gramas por metro quadrado). Isso pode indicar que o modelo enfrentou dificuldades em aprender os padrões relacionados a esta variável, devido à escassez de dados de treinamento nessa faixa específica, e por isso, a quantidade de dados disponíveis para o aprendizado não foi suficiente para capturar os comportamentos complexos associados a altas gramaturas.

Uma observação importante a ser destacada é que o volume de produção de papéis de imprimir e escrever com alta gramatura geralmente é significativamente menor em comparação aos papéis de baixa gramatura. Isso ocorre devido a várias razões, sendo uma delas a demanda de mercado.

O gráfico mostrado na Figura 4.2 é uma representação das métricas de perda durante o treinamento do modelo, indicando o quão distantes as previsões do modelo estão dos valores reais. Ao analisar o gráfico, podemos observar a tendência de diminuição da perda tanto no treinamento quanto na validação ao longo das épocas. Isso

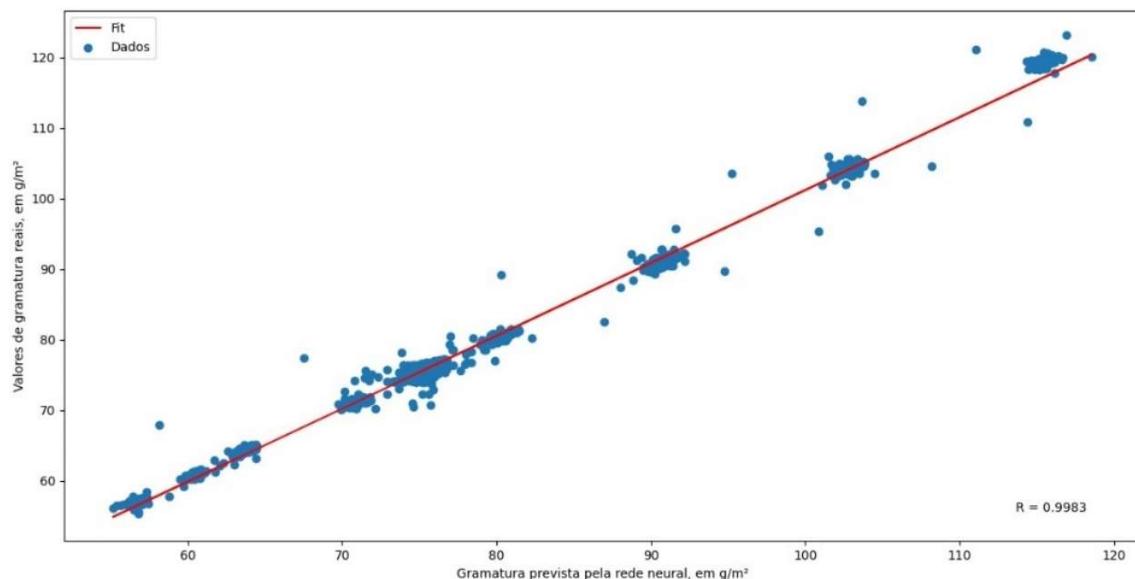
significa que o modelo está aprendendo a fazer previsões mais precisas à medida que o treinamento avança.

Figura 4.2 – Perda durante o treinamento e teste do modelo



A redução da perda é um indicativo de que a rede neural está se ajustando aos dados de treinamento e se tornando mais capaz de expandir a base de dados. A proximidade entre a perda do treinamento e a perda da validação é um bom sinal, indicando que o modelo não está sofrendo de *overfitting*; isso ocorre quando o modelo se ajusta muito bem aos dados de treinamento, mas tem dificuldade em generalizar para novos dados.

Figura 4.3 – Análise de regressão para os dados usados na construção da rede neural

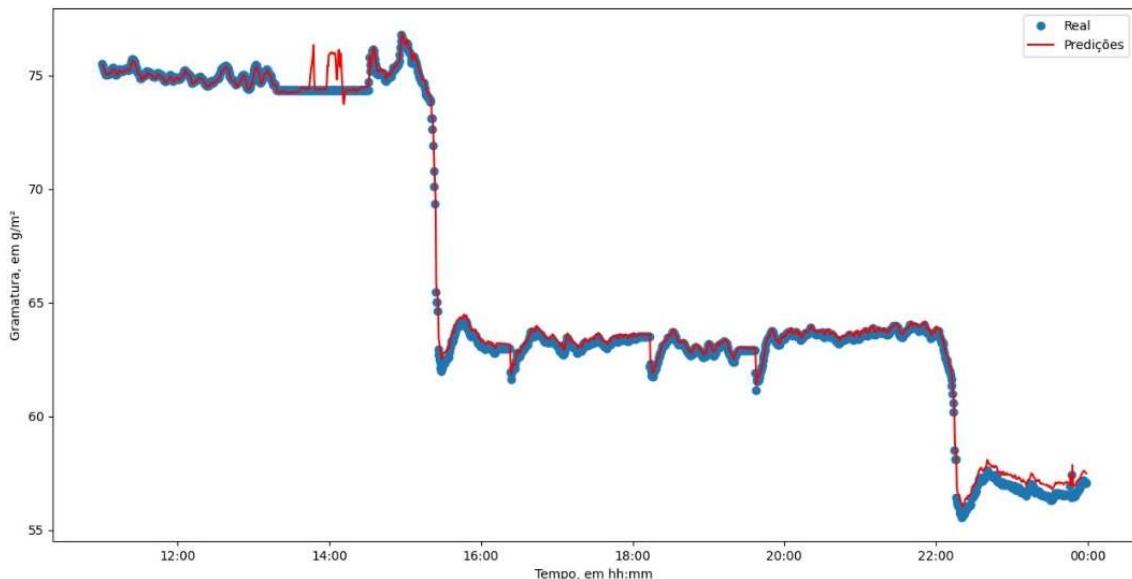


O gráfico acima representa a relação estreita entre os valores previstos e os valores reais por meio de um ajuste linear. Os pontos dispersos representam os pares, na qual para cada ponto indica um valor previsto e o valor real correspondente; dessa forma, podemos visualizar a proximidade entre as estimativas e os valores reais. A linha de ajuste linear, determinada pelos coeficientes de regressão calculados previamente, mostra a tendência geral dos dados e como as previsões variam em relação ao seu valor *target*.

O valor de R, visto no gráfico, conclui que existe uma correlação forte entre os valores reais de gramatura e os gerados pela rede neural. Em outras palavras, o modelo foi capaz de capturar de forma precisa a relação entre as variáveis e estimar resultados com alta precisão.

Uma nova base de dados, que não foi utilizada durante a etapa de treinamento do modelo, foi usada para validar novamente a rede neural, proporcionando um conjunto de exemplos independentes para avaliar a capacidade de se adaptar a diferentes padrões e variações nos dados de entrada. Ao utilizar uma nova base de dados que contém duas trocas de gramatura durante o dia, estamos simulando condições reais para observar como o modelo reage ao lidar com mudanças abruptas e dados contínuos, levando em consideração a natureza dinâmica do problema e fornecendo ao modelo um desafio adicional em termos de previsão precisa da gramatura.

Figura 4.4 – Análise comparativa entre valores reais e previsões de gramatura



Um pouco antes da primeira troca de gramatura pode-se observar um período de quebra de folha, caracterizado pela parte constante do gráfico. A quebra de folha é um

problema comum em processos de fabricação de papel contínuos, onde a folha se separa em duas ou mais partes indesejadas durante a produção. Isso pode ser causado por tensão excessiva aplicada ao papel, umidade inadequada, problemas na formação da folha, defeitos (furos, pingos de água, defeitos de borda, manchas), entre outros.

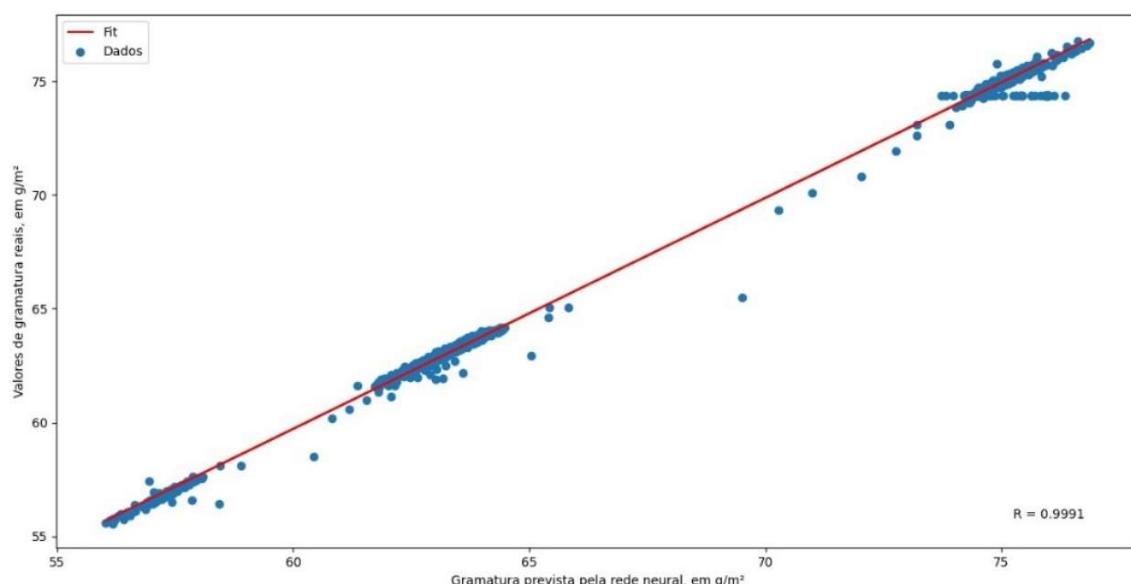
Durante a ocorrência de quebra, é comum que o sensor responsável pela medição registre o último valor medido de forma constante, pois a folha de papel foi interrompida e não está mais passando pelo sensor, resultando na manutenção do valor anteriormente registrado.

Além disso, diversas medidas são tomadas para restabelecer a operação normal. Essas medidas incluem ajustes nas variáveis do processo, como velocidade, tensão, temperatura e outros parâmetros relevantes, que podem resultar em perturbações nas variáveis de entrada da rede neural utilizada para prever a gramatura do papel, o que explica o comportamento do modelo nessa faixa.

É importante mencionar também que, ao trocar para uma gramatura de 56 g/m² (gramas por metro quadrado), o comportamento do modelo pode ter ficado um pouco fora do esperado. Isso pode ser atribuído ao fato de que a transição para uma gramatura mais baixa pode marcar o início do processo de fabricação do papel, no qual são feitos vários ajustes e variações para atingir o *set point*.

Essa redução na espessura e densidade pode afetar o comportamento do processo de fabricação, uma vez que requer um manuseio mais cuidadoso e uma maior precisão nos ajustes dos equipamentos.

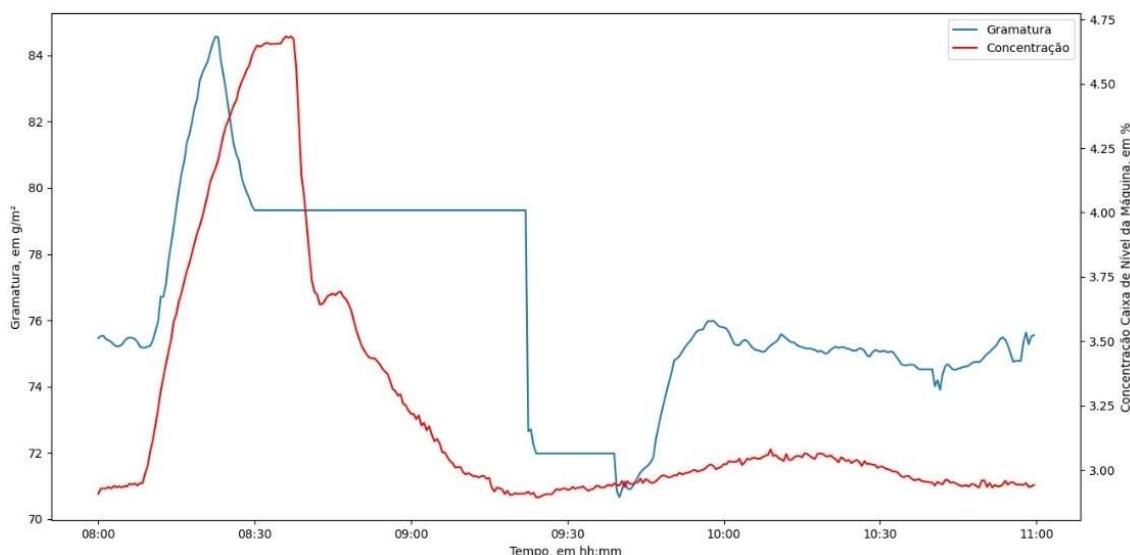
Figura 4.5 – Análise de regressão para os dados usados na análise comparativa



Ao observar o gráfico de regressão, é possível visualizar como os pontos dispersos se aproximam da linha de tendência, e pode-se observar que o valor do coeficiente de correlação continua sendo alto, reforçando a confiabilidade do modelo para previsão de baixas gramaturas. Essa consistência na performance, mesmo quando aplicado a dados não vistos anteriormente, é um indicativo da robustez e generalização da rede neural.

Uma abordagem utilizada para testar e aprimorar a previsão de gramatura foi a utilização de uma ocorrência real de variação durante o processo de fabricação do papel. Essa variação ocorreu de forma não planejada, representando um desafio adicional na capacidade do modelo de lidar com eventos inesperados e de adaptar-se a mudanças nas condições de produção, como instabilidades no processo, mudanças nas matérias-primas, ajustes inadequados dos equipamentos etc.

Figura 4.6 – Comportamento da gramatura durante ocorrência de variação, em 23/09/22



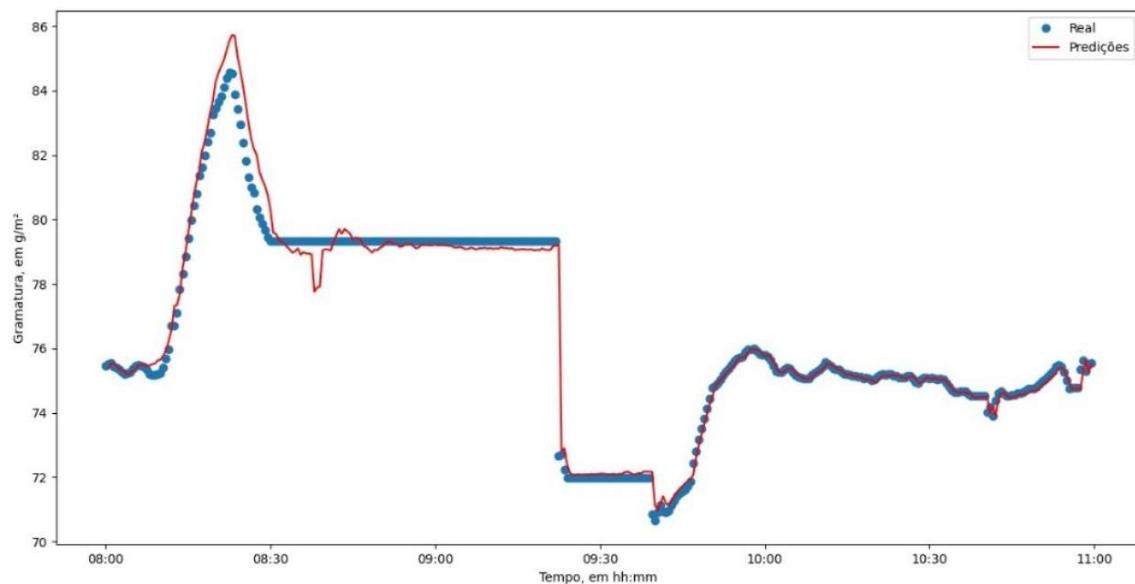
Em setembro de 2022, uma instância de instabilidade de gramatura ocorreu devido parada dos engrossadores para manutenção, interrompendo o fornecimento de filtrado para o tanque 2. Como a alimentação de água fresca não consegue regular sozinha o nível do tanque, essa perturbação afetou também o nível do tanque 24, responsável pelo controle de concentração da massa em pontos cruciais. Como resultado, a bomba de entrada da máquina não conseguiu operar adequadamente devido a ultrapassagem do seu range de trabalhabilidade.

É importante ressaltar que não ocorre descontinuidade nos dados de gramatura, como pode parecer de primeira instância. A queda brusca do valor pode ser explicada pelo fenômeno de quebra de folha.

Por volta das 8:29h, a concentração da massa na caixa de nível da máquina ultrapassou a marca de 4,5%, perturbando a alimentação do processo e, por fim, zerando o avanço de massa. Durante o período de restabelecimento do fluxo de processo (passagem de ponta da folha), o sensor de gramatura registrou o último valor medido, enquanto que outras variáveis como a taxa de produção foram zeradas, devido o incidente de descontinuidade da passagem da folha pelo sensor.

Pouco depois das 9:20h, a concentração da caixa começou a estabilizar, e os operadores conseguiram passar a ponta da folha pela máquina. Nesse momento, os sensores começaram a aferir novamente as propriedades do sistema, saltando do último valor gerado para o novo valor medido em questão de segundos, causando o efeito visto no gráfico. Logo em seguida, a folha quebrou novamente dois minutos depois e a passagem de ponta só foi possível após o controle da concentração da caixa de nível para o seu valor padrão, entre 2,9~3,1%.

Figura 4.7 – Análise comparativa durante ocorrência de variação de gramatura



O modelo se adaptou de forma eficiente e conseguiu reproduzir o comportamento da gramatura durante o evento não previsto. Foi observado que o *soft sensor* foi capaz de prever com antecedência de sete minutos o aumento da gramatura, e essa capacidade tem relação direta com as variáveis de entrada utilizadas.

Ao considerar variáveis chaves, a rede neural é capaz de capturar as interações e correlações entre elas e a gramatura do papel; dessa forma, quando ocorre uma variação em uma ou mais dessas variáveis, o modelo consegue antecipar o efeito dessa mudança.

Essa capacidade de prever antes do acontecimento se deve ao aprendizado das relações complexas entre as variáveis ao longo do treinamento da rede. Assim, o modelo se torna uma ferramenta valiosa para identificar potenciais problemas e tomar medidas proativas para evitar desvios indesejados no processo de produção.

5 CONCLUSÕES

O comportamento do modelo desenvolvido para baixas e altas gramaturas apresentou características distintas, revelando nuances importantes na sua capacidade de previsão em diferentes faixas de valores.

Esses resultados indicam que o desempenho é influenciado pelo intervalo de gramaturas considerado. O modelo tem uma tendência natural de ter um melhor desempenho em faixas de baixas gramaturas, onde a precisão nas previsões se manteve consistente, com erros médios e quadráticos médios reduzidos. No entanto, à medida que a gramatura aumenta, a complexidade do processo industrial e a variabilidade associada a essas condições podem impactar a precisão das estimativas.

Uma recomendação importante para melhorar as aproximações do *soft sensor* em relação às altas gramaturas do papel é a utilização de uma base de dados maior e mais extensiva na etapa de treinamento. A quantidade e diversidade dos dados de treinamento são fatores cruciais para o desempenho e generalização da rede neural LSTM.

Ao incorporar uma base de dados mais abrangente, que inclua uma ampla faixa de gramaturas, é possível fornecer ao modelo informações mais completas e representativas do processo produtivo. Isso permite que o *soft sensor* aprenda com uma variedade maior de padrões e relações presentes nos dados, resultando em estimativas mais precisas para as altas gramaturas. Embora a produção de altas gramaturas possa ser menos frequente, a coleta de dados durante um período mais longo pode compensar essa limitação, garantindo um número adequado de exemplos para treinamento do modelo.

Em suma, os resultados obtidos neste estudo forneceram uma base sólida para as conclusões em relação aos objetivos estabelecidos. Eles demonstraram a relevância das técnicas de *soft sensors* baseados em redes neurais LSTM na modelagem e previsão de variáveis em processos industriais, ressaltaram a importância do pré-processamento cuidadoso dos dados e validaram a acurácia do modelo desenvolvido. Essas conclusões reforçam o potencial de aplicação dessas técnicas na indústria de fabricação de papel, oferecendo benefícios como o controle otimizado do processo produtivo, a melhoria da

qualidade do produto final e a otimização dos recursos utilizados. No entanto, é importante destacar a necessidade de aprimoramento contínuo do modelo, coleta de mais dados e adaptação a condições de produção variáveis para maximizar seu desempenho e aplicabilidade em cenários industriais.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKESSON, J.; SLATTEKE, O. Modeling, Calibration and Control of a Paper Machine Dryer Section. In: 5th International Modelica Conference - September 4th - 5th, 2006, Vienna. *Anais...* Vienna: Modelica Association, 2006.
- ALVES, F.; FIORIO, V. *Como é fabricado o papel?* Industria Hoje, 25 dez. 2014. Disponível em: <https://industriahoje.com.br/como-e-fabricado-o-papel>. Acesso em: 12 mai. 2023.
- ANDRIDGE, R.R.; LITTLE, R.J.A. A review of hot deck imputation for survey nonresponse. *International Statistical Review*, vol. 78, no. 1, pp. 40-64, 2010.
- ARAÚJO NETO, A. P. de. *Redes neurais artificiais aplicadas à mudança de especificação de etanol via destilação extractiva*. Trabalho de Conclusão de Curso (Bacharel em Engenharia Química). Universidade Federal de Campina Grande, Campina Grande, PB, p. 47. 2018.
- BACHMEIER, J. *Voith's SpeedSizer impresses customers all over the world*. 2014. Disponível em: <https://voith.com/corp-en/news-room/press-releases-54167.html>. Acesso em: 12 mai. 2023.
- BRAGA, A. de.; LUDERMIR, T. B.; CARVALHO, A. C. P. de L. F. *Redes Neurais Artificiais teoria e aplicações*, Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A., 2000.
- CAMPOS, E. S. Curso básico de fabricação de celulose e papel. *Eucalyptus Online Book & Newsletter*, 2011.
- CASTELÃO, R. *Utilização de Redes Neurais para Previsões do mercado de ações*. São Paulo: Unicamp, 2018. 30 p. (Relatório Técnico - IC-PFG-18-01).
- COMMISSION, E. *Integrated Pollution Prevention and Control (IPPC): Reference Document on Best Available Techniques in the Pulp and Paper Industry*. 2001.
- D'ALMEIDA, M. L. O.; KOGA, M. E. T.; FERREIRA, D. C.; PIGOZZO, R. J. B.; TOUCINI, R.; REIS, H. M. dos; VIANA, E. F. *Celulose SENAI*: São Paulo, 2013.
- DAVIES, L.; GATHER, U. The identification of multiple outliers. *Journal of the American Statistical Association*, vol. 8, no. 423, pp. 782-792, 1993.

FORTUNA, L.; GRAZIANI, S.; RIZZO, A.; XIBILIA, M. G. *Soft sensors for monitoring and control of industrial processes*. 1. ed. Springer, 2006.

GOMAA, W. H. The Impact of Deep Learning Techniques on SMS Spam Filtering. *International Journal of Advanced Computer Science and Applications*, v. 11, n. 1, p. 544-549. Jan./Feb. 2020.

HAMAGUCHI, M. *Análise do circuito de água em processo de fabricação de papel imprensa integrada com produção de pastas termomecânicas*. 2007. Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2007.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning. Springer Series in Statistics*. Springer, New York, 2001.

HAYKIN, S. *Redes Neurais: Princípios e prática*. 2. ed. Tradução Paulo Martins Engel. Porto Alegre: Bookman, 2001.

HENTINEN, K. Modelling the dewatering in the forming section of a paper machine. 2010. Master's Thesis - University of Eastern Finland, Kuopio, 2010.

HUNTER, D. *Papermaking: the history and technique of an ancient craft*. 2. ed. New York: Dover Publications, INC., 1978.

JUDASZ, A. *Estimation of impact of alternative papermaking additives on paper web dewatering intensity and paper properties*. 2009. Dissertação de Mestrado - Faculdade de Química, Universidade Técnica de Lodz, 2009.

KADLEC, P.; GABRYS, B.; STRANDT, S. Data-driven soft sensors in the process industry. *Comput. Chem. Eng.*, v. 33, n. 4, p. 795–814, 2009.

KADLEC, P.; GABRYS, B. Local learning-based adaptive soft sensor for catalyst activation prediction. *AICHE Journal*, v. 57, n. 5, p. 1288-1301, 2011.

LINDSTRÖM, S. *Modelling and simulation of paper structure development*. 2008. Master's Thesis - Mid Sweden University, Sundsvall, 2008.

LIU, H.; SHAH, S.; JIANG, W. On-line outlier detection and data cleaning. *Computers and Chemical Engineering*, vol. 28, no. 9, pp. 1635-1647, 2004.

LIU, J.; GONG, X. Attention mechanism enhanced LSTM with residual architecture and its application for protein-protein interaction residue pairs prediction. *BMC Bioinformatics*, v. 20, n. 1, 2019.

LU, N.; YANG, Y.; GAO, F.; WANG, F. Multirate dynamic inferential modeling for multivariable processes. *Chemical Engineering Science*, v. 59, n. 4, p. 855-864, 2004.

MENDES, A. H. T. *Higroexpansibilidade de papel reprográfico produzido com fibras de eucalipto em máquina industrial*. 2006. Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2006.

MONTGOMERY, D. C.; RUNGER, G. C. *Applied statistics and probability for engineers*. John Wiley & Sons, 2010.

MORAES, C. C.; PESCIO, P. H. Otimização do controle de velocidade em bomba de mistura na caixa de entrada para processo de fabricação de papel. In: 2010 9th IEEE/IAS International Conference on Industry Applications - INDUSCON 2010, São Paulo. *Anais...* São Paulo: IEEE, 2010.

MORAIS, E. C. *Reconhecimento de padrões e redes neurais artificiais em predição de estruturas secundárias de proteínas*. Tese (Doutorado em Engenharia de Sistemas e Computação). Universidade Federal do Rio de Janeiro, Rio de Janeiro, p. 148. 2010.

NELLES, O. *Nonlinear System Identification - From Classical Approaches to Neural Networks and Fuzzy Models*. Springer, 2001.

NUNES, D. C. L. *Conservação de água em máquina de fabricação de papel: o caso da Bahia Sul Papel e Celulose S.A.* 2007. Dissertação (Mestrado) - Universidade Federal de Itajubá, Itajubá, 2007.

PEARSON, R.K. Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology*, vol. 10, no. 1, pp. 55-63, 2002.

PESCIO, Paulo Henrique. *Análise da integração da tecnologia da automação aplicada ao processo de fabricação de papel*. 2016. Dissertação (Mestrado em Sistemas de Potência) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2016.

PHILIPP, P.; D'ALMEIDA, M. L. O. *Celulose e papel – Volume I – Tecnologia de fabricação da pasta celulósica*. 2. ed. São Paulo: IPT - Instituto de Pesquisas Tecnológica do Estado de São Paulo - Centro Técnico em Celulose e Papel, 1988.

PIOTTO, Z. C. *Eco-eficiência na Indústria de Celulose e Papel – Estudo de Caso*. Universidade de São Paulo, 2003.

ROBUSTI, C.; VIANA, E. F.; JÚNIOR, F. F.; GOMES, I.; TOGNETTA, L.; SANTOS, O. dos; DRAGONI, P. *Papel*. 1. ed. São Paulo: SENAI-SP, 2014.

RODRIGUES, N. T. J. *Redes neurais artificiais LSTM aplicadas na previsão de padrões de um forno de reaquecimento de tarugos*. Dissertação (Mestrado em Engenharia Química) - Universidade Federal de Campina Grande, Centro de Ciências e Tecnologia, Programa de Pós-Graduação em Engenharia Química, Campina Grande, PB, 2022.

SANTOS, C. P.; NUNES, I.; EDUARDO, J.; MOREIRA, B.; BRASILEIRO, B. Como se fabrica papel. *Química Nova na Escola*, v. 14, n. 1, 2001.

SÉKULA, P. R. *Uso de modelo de referência para melhoria do processo de fabricação de papel kraft*. 2011. Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2011.

SHREVE, R. N.; BRINK JR., J. A. *Indústrias de Processos Químicos*. 4. ed. Rio de Janeiro: Guanabara Dois, 1980.

SOUZA, F. A. A.; ARAÚJO, R.; MENDES, J. Review of soft sensor methods for regression applications. *Chemometrics and Intelligent Laboratory Systems*, v. 152, p. 69-79, 2016.

STAUDEMAYER, R. C.; MORRIS, E. R. *Understanding LSTM--a tutorial into long short-term memory recurrent neural networks*, 2019.

TEIXEIRA, M. B. D.; OLIVEIRA, R. A.; GATTI, T. H.; SUAREZ, P. A. Z. O Papel: Uma Breve Revisão Histórica, Descrição da Tecnologia Industrial de Produção e Experimentos para Obtenção de Folhas Artesanais. *Revista Virtual de Química*, v. 9, n. 3, p. 1364-1380, jun. 2017.

TSYMBAL, A. *The Problem of Concept Drift: Definitions and Related Work*. Technical Report. Department of Computer Science, Trinity College, Dublin, Ireland, 2004.

VALMET. *Cilindro de Secagem Valmet*: Máxima eficiência com mínimo consumo de energia, 2023. Disponível em: <https://www.valmet.com/pt/cartao-e-papel/maquinas-de-cartao-e-papel/cilindro-de-secagem/>. Acesso em: 12 mai. 2023.

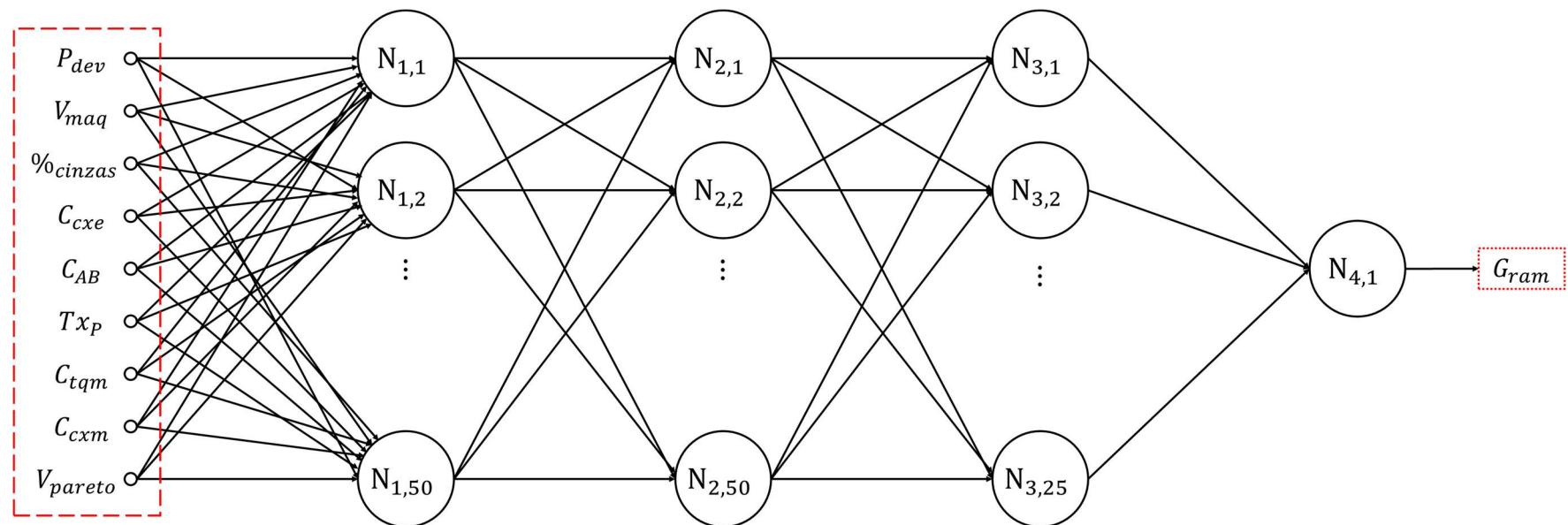
VOITH. *Schoellershammer PM 6 gets off to a great start: Voith's XcelLine and BlueLine impress*, 2017. Disponível em: <https://voith.com/corp-en/industry-solutions/papermaking/nextlevel/article-75061.html>. Acesso em: 12 mai. 2023.

WARNE, K.; PRASAD, G.; REZVANI, S.; MAGUIRE, L. Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. *Engenharia de Aplicações da Inteligência Artificial*, v. 17, n. 8, p. 871-885, 2004.

WU, Y.; LUO, X. A novel calibration approach of soft sensor based on multirate data fusion technology. *Journal of Process Control*, v. 20, n. 10, p. 1252-1260, 2010.

XIE, L.; YANG, H.; HUANG, B. Fir model identification of multirate processes with random delays using EM algorithm. *AIChE Journal*, v. 59, n. 11, p. 4124-4132, 2013.

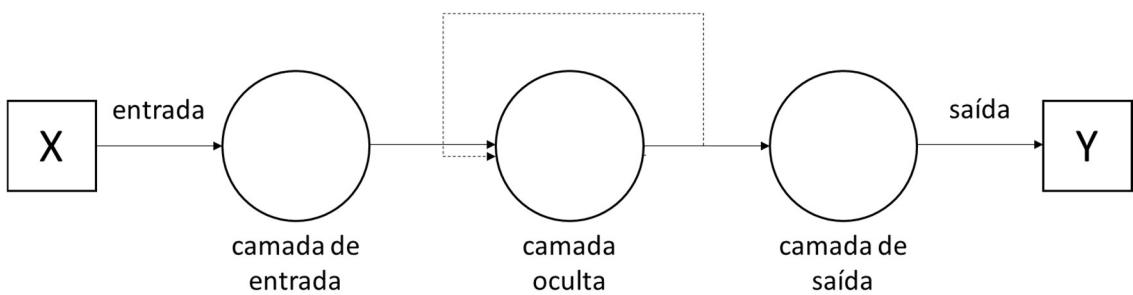
APÊNDICE A – TOPOLOGIA DA REDE NEURAL ARTIFICIAL PROPOSTA NO TRABALHO



ANEXO A – MODELO MATEMÁTICO DA RNN LSTM EXTRAÍDO DE LIU & GONG, 2019

Rede Neural Recorrente (RNN), proposta por *Hochreiter*, é uma das principais redes neurais no campo do aprendizado profundo, que atua como uma ponte para conectar a informação do passado ao presente. Ela é baseada no algoritmo de retropropagação e leva em consideração o fator temporal, sendo assim, a RNN é uma forma de algoritmo de retropropagação no tempo (BPTT). Além disso, ela é capaz de lidar com dados sequenciais, incluindo dados temporais e espaciais, devido às suas propriedades.

Fig. 1. Modelo padrão de uma RNN



Fonte: Adaptado de LIU & GONG, 2019

Observe a Fig. 1 da RNN padrão, onde a informação é propagada de forma direta dos *inputs* para os *outputs*. Podemos descrever esse fluxo de informação por meio de uma série de equações. x denota o valor do vetor de entrada, x_i^t denota o valor do i -ésimo *input* do vetor x no instante de tempo t , e w_{ij} denota o peso da unidade i para a unidade j . Para a unidade da camada oculta h , denotamos a entrada da unidade da camada oculta h no instante de tempo t :

$$a_h^t = \sum_{i=1}^I w_{ih} x_i^t + \sum_{h'=1}^H w_{h'h} b_{h'}^{t-1} \quad (1)$$

A saída da unidade da camada oculta h no instante de tempo t é representada como b_h^t , e a função de ativação é θ_h , então:

$$b_h^t = \theta(a_h) \quad (2)$$

A entrada da camada de saída pode ser calculada ao mesmo tempo:

$$a_k^t = \sum_{h=1}^H w_{hk} b_h^t \quad (3)$$

Assim como o algoritmo de retropropagação padrão, BPTT também é uma aplicação repetida da regra da cadeia. Para os gradientes das funções de perda em RNN, a influência da função de perda no oculto não ocorre apenas através da saída da camada oculta, mas também por meio do seu próximo passo de tempo:

$$\delta_h^t = \theta'(a_h^t) \left(\sum_{k=1}^K \delta_k^t w_{hk} + \sum_{h'=1}^{t+1} w_{hh'} \right), \quad (4)$$

onde

$$\delta_j^t \stackrel{\text{def}}{=} \frac{\partial \mathcal{L}}{\partial a_j^t} \quad (5)$$

Então podemos obter a derivada do peso de toda a rede, respectivamente:

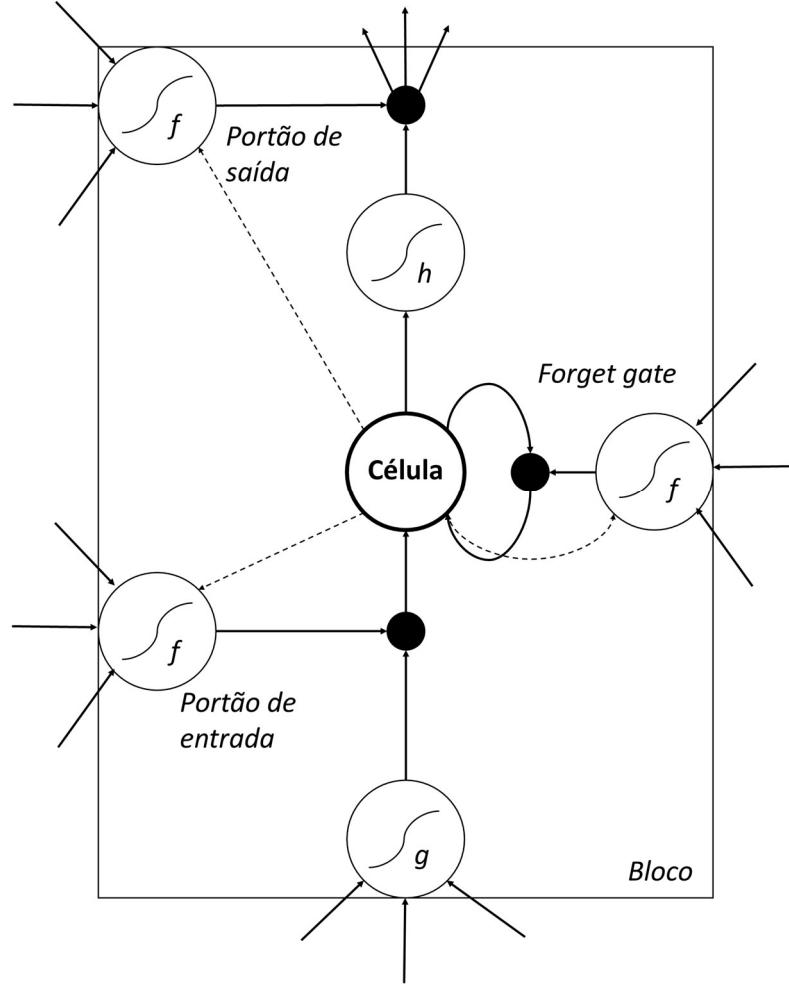
$$\frac{\partial \mathcal{L}}{\partial w_{ij}} = \sum_{t=1}^T \frac{\partial \mathcal{L}}{\partial a_j^t} \frac{\partial a_j^t}{\partial w_{ij}} = \sum_{t=1}^T \delta_j^t b_i^t \quad (6)$$

Long Short-Term Memory, como uma variante da RNN, proposta por *Hochreiter* e mostrada na Figura 2, consiste em um bloco que possui três portões (portão de entrada, de esquecimento e de saída), cujas probabilidades de ativação variam de 0 (o portão está fechado) a 1 (o portão está aberto), e algumas células que podem lembrar informações e transmiti-las para o próximo passo, enquanto a unidade de camada oculta na RNN é substituída por três portões. Os valores de saída do portão de entrada e do portão de esquecimento são determinados pelos estados anteriores das células e pelos valores de entrada.

Os subscritos i , ϕ e ω denotam os portões de entrada, esquecimento e saída do bloco, respectivamente, e c denota uma das células de memória C . O peso de visão direta (*peephole*) da célula c para os portões de entrada, esquecimento e saída é denotado como w_{ci} , $w_{c\phi}$ e $w_{c\omega}$, respectivamente. s_c^t denota o estado da célula c no tempo t . f , g e h são as funções de ativação dos portões, entrada da célula e saída, respectivamente. Vamos

denotar I como o número de entradas, K como o número de saídas e H como o número de células na camada oculta.

Fig. 2. Bloco de memória com uma célula da rede neural LSTM



Fonte: Adaptado de LIU & GONG, 2019

Ao observar o *framework* apresentado na Fig. 2, podemos obter as equações:
portão de entrada

$$a_t^t = \sum_{i=1}^I w_{it} x_i^t + \sum_{h=1}^H w_{ht} b_h^{t-1} + \sum_{c=1}^C w_{ct} s_c^{t-1}, \quad (7)$$

$$b_t^t = f(a_t^t), \quad (8)$$

portão de esquecimento

$$a_\phi^t = \sum_{i=1}^I w_{i\phi} x_i^t + \sum_{h=1}^H w_{h\phi} b_h^{t-1} + \sum_{c=1}^C w_{c\phi} s_c^{t-1}, \quad (9)$$

$$b_\phi^t = f(a_\phi^t), \quad (10)$$

célula

$$a_c^t = \sum_{i=1}^I w_{ic} x_i^t + \sum_{h=1}^H w_{hc} b_h^{t-1}, \quad (11)$$

$$s_c^t = b_\phi^t s_c^{t-1} + b_l^t g(a_c^t), \quad (12)$$

portão de saída

$$a_\omega^t = \sum_{i=1}^I w_{i\omega} x_i^t + \sum_{h=1}^H w_{h\omega} b_h^{t-1} + \sum_{c=1}^C w_{c\omega} s_c^{t-1}, \quad (13)$$

$$b_\omega^t = f(a_\omega^t), \quad (14)$$

saída da célula

$$b_c^t = b_\omega^t h(s_c^t) \quad (15)$$

Comparada com outras RNN, LSTM é mais fácil de alterar dinamicamente o peso do modelo autorrecursivo adicionando as *gates*, e lidar com dados de diferentes escalas com melhor desempenho.