

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Hesloan dos Santos Viana

**Criação de um sistema de recomendação de
fornecedores para restaurantes utilizando
mineração de texto**

Uberlândia, Brasil

2023

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Hesloan dos Santos Viana

**Criação de um sistema de recomendação de fornecedores
para restaurantes utilizando mineração de texto**

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Sistemas de Informação.

Orientador: Elaine Ribeiro de Faria Paiva

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2023

Hesloan dos Santos Viana

Criação de um sistema de recomendação de fornecedores para restaurantes utilizando mineração de texto

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Sistemas de Informação.

Trabalho aprovado. Uberlândia, Brasil, 21 de novembro de 2023:

Prof^ª. Dr^ª. Elaine Ribeiro de Faria
Orientador

Prof. Dr. Ivan da Silva Sendin

**Prof. Dr. João Henrique de Souza
Pereira**

Uberlândia, Brasil
2023

Agradecimentos

O meu primeiro agradecimento vai à minha mãe e irmã, duas incríveis pessoas que me ajudam a todo custo e a todo momento da minha vida, que me inspiraram desde criança para que eu desafiasse todas as dificuldades da vida com a noção de que elas vão estar lá para me ajudar.

Agradeço também à minha esposa, que por anos me apoiou em todas as formas, que viu de perto todo o esforço e dificuldade que tive durante o meu tempo na faculdade, mas que sempre estava lá para reforçar que eu conseguiria superar este desafio, por muitas vezes confiando em mim mais do que eu mesmo.

Agradeço também aos meus amigos, que por todos esses anos estiveram ao meu lado, me dando forças para não desistir, e com quem, por diversas vezes, tiveram que ouvir desabaços sobre as dificuldades da faculdade.

Por fim, agradeço a todos os professores da FACOM, e em especial à Professora Dra. Elaine Ribeiro de Faria, por todos os ensinamentos compartilhados durante a faculdade e por todo o apoio durante este projeto.

Resumo

Durante a pandemia da COVID-19, os restaurantes mudaram o seu estilo de funcionamento, e, aproveitando que os aplicativos de *delivery* tiveram uma expansão rápida entre os usuários durante esse período (DATA.AI, 2021), surgiram diversos novos restaurantes com o intuito de trabalharem apenas via *delivery*. Apesar de estar em um aplicativo de *delivery* facilitar para os restaurantes que os consumidores tenham acesso aos seus pratos, os problemas comuns de gerência de estabelecimentos desse ramo ainda continuam existindo, como é o caso do problema de gastos e controle de estoque de insumos para os estabelecimentos.

Aproveitando que possui a liderança no meio de aplicativos de *delivery* no Brasil, o *iFood* criou uma nova plataforma, chamada de *iFood Shop*, com o objetivo de conectar fornecedores de insumos, como pequenos mercados e grandes atacadões, com os restaurantes cadastrados em sua base. Desta forma, um restaurante conseguiria vender seus pratos para os consumidores, via aplicativo de *delivery* do *iFood*, e comprar os ingredientes necessários para o seu estoque, via sistema do *iFood Shop*.

Considerando este cenário, este projeto surgiu com o intuito de ajudar o restaurante nesta tarefa de compra de insumos. É realizado um tratamento sobre os pratos que os estabelecimentos possuem cadastrados no aplicativo de *delivery* e, a partir dos resultados destes tratamentos, são identificados os ingredientes que os restaurantes utilizam em seus pratos para então realizar a recomendação dos melhores fornecedores possíveis a ele, baseando-se em um cálculo do quanto o fornecedor consegue ofertar do total de ingredientes daquele restaurante.

Em média, 87% dos ingredientes dos restaurantes foram ofertados pelo melhor fornecedor encontrado para eles, o que garante um número satisfatório para que os estabelecimentos consigam comprar a maior parte dos insumos necessários para o seu funcionamento de um único fornecedor, facilitando na complexa atividade de compra de insumos, um dos principais motivos que levam restaurantes à falência.

Palavras-chave: Mineração de texto, pré-processamento de dados, aplicativos de *delivery*, sistema de recomendação.

Lista de ilustrações

Figura 1 – Extração dos pratos dos restaurantes	29
Figura 2 – Um prato cadastrado no aplicativo	39
Figura 3 – Um insumo do fornecedor cadastrado no aplicativo	40

Lista de tabelas

Tabela 1 – Tabela com trabalhos relacionados	25
Tabela 2 – Colunas da base inicial de restaurantes e o que cada uma representa . .	31
Tabela 3 – Colunas da base inicial de fornecedores e o que cada uma representa .	31
Tabela 4 – Tabela demonstrando concatenação de palavras	34
Tabela 5 – Tabela final dos fornecedores	36
Tabela 6 – Tabela final dos restaurantes	36
Tabela 7 – Tabela demonstrando tratativa nos nomes das cidades	39
Tabela 8 – Prato de exemplo após o pré processamento inicial	39
Tabela 9 – Volumetria da base de restaurantes	39
Tabela 10 – Volumetria e principais nomes de pratos presentes por tipo de categoria do restaurante	40
Tabela 11 – Volumetria da base de fornecedores	41
Tabela 12 – Exemplos de descrições de pratos após a remoção de <i>stopwords</i>	41
Tabela 13 – Tabela com palavras frequentes que foram removidas da base de for- necedores	42
Tabela 14 – Contagem de frequência de palavras na base	42
Tabela 15 – Exemplos de descrições de pratos após a remoção de <i>stopwords</i> adicionais	43
Tabela 16 – Exemplos de concatenação de palavras feitas	44
Tabela 17 – Contagem de vezes em que as palavras óleo, manteiga e azeite aparecem antes e depois do pré-processamento realizado	44
Tabela 18 – Transformação de pratos em ingredientes específicos	45
Tabela 19 – Exemplos de cardápios de três restaurantes distintos antes e depois dos pré processamentos.	46
Tabela 20 – Volumetria e principais nomes de pratos presentes por tipo de categoria do restaurante	47

Lista de abreviaturas e siglas

CRM	Customer Relationship Management
SQL	Structured Query Language
ID	Identification
KPI	Key Performance Indicator

Sumário

1	INTRODUÇÃO	10
1.1	Justificativa	11
1.2	Objetivos	13
1.2.1	Objetivos Específicos	13
1.3	Organização do Trabalho	13
2	REVISÃO BIBLIOGRÁFICA	15
2.1	Lojas Virtuais	15
2.2	<i>Marketplace</i>	16
2.3	<i>Food Delivery</i>	16
2.4	Pandemia e <i>Dark Kitchens</i>	18
2.5	Dificuldades de restaurantes	19
2.6	Mineração de Texto	20
2.7	Etapas da Extração de Conhecimento a partir de Textos	21
2.7.1	Coleta de dados	21
2.7.2	Pré-Processamento	22
2.7.3	Mineração de Dados	22
2.8	Sistemas de recomendação	23
2.9	Trabalhos relacionados	24
2.10	Considerações Finais	27
3	DESENVOLVIMENTO	28
3.1	Visão Geral da Proposta	28
3.2	Coleta e pré-processamento Inicial	29
3.2.1	Base dos restaurantes	29
3.2.2	Base dos fornecedores	31
3.3	Pré-processamento da base textual	32
3.3.1	Base dos restaurantes	32
3.3.2	Base dos fornecedores	32
3.4	Transformação de pratos em ingredientes	33
3.5	Agrupamento de ingredientes e insumos	35
3.6	Sistema de recomendação	36
3.7	Considerações Finais	36
4	RESULTADOS	38
4.1	Coleta de dados e volumetria da base	38

4.1.1	Base dos restaurantes	38
4.1.2	Base dos fornecedores	40
4.2	Resultados pré-processamento textual	41
4.3	Resultados da transformação de pratos em ingredientes	42
4.4	Recomendação de fornecedores	45
4.5	Considerações finais	47
5	CONCLUSÃO	48
5.1	Contribuições	48
5.2	Trabalhos futuros	49
	REFERÊNCIAS	50

1 Introdução

O conceito de compras *online* surgiu em 1979 a partir do momento em que o empresário e inventor inglês Michael Aldrich conseguiu realizar a conexão de sua televisão modificada com um computador que estava presente em outro local a partir de uma linha telefônica (MIVA, 2020). O sucesso desse teste permitiu que diversas empresas, dos mais variados segmentos possíveis, pudessem vender seus produtos para o consumidor final de uma forma mais simples e direta, não sendo mais necessário um contato físico entre o cliente e o estabelecimento.

Em 1992 foi criada uma das primeiras empresas focadas em venda *online*, a *Book Stacks Unlimited* (VYSHNAVI; MALLIKA, 2020) focada na venda de livros para seus consumidores. Logo após a criação de uma livraria *online* surgiu a *Amazon*, em 1994, que é considerada hoje a maior empresa de *e-commerce* do mundo (Mohammed Saqib, 2023). Também em 1994, a *Pizza Hut*, uma das maiores redes de pizza dos Estados Unidos, iniciou um teste apresentando para seus clientes a *PizzaNet*, uma nova forma de fazer os seus pedidos que não seja por telefone ou no próprio restaurante, ou seja, o cliente poderia agora fazer seus pedidos de forma *online* (MICHAEL SCHRAGE, 1994).

Com o destaque da *Amazon* nesse cenário de vendas *online*, o conceito de *marketplaces* passou a ficar em alta. *Marketplaces* são plataformas que reúnem diversos produtos de diversas lojas diferentes dando aos usuários dessas plataformas uma gama gigantesca de variedade (AMAZON, 2021). Assim, o consumidor não precisará procurar por *sites* específicos para cada tipo de mercadoria que o mesmo queira adquirir.

O primeiro *marketplace* focado em restaurantes foi o *World Wide Waiter*, que foi criado em 1995 reunindo 60 restaurantes diferentes e fazendo dentro da própria plataforma uma categorização por tipo de cozinha e por proximidade com a localização do cliente (WAITER, 2023). Nos dias atuais existe uma competitividade enorme nesse ramo de *marketplace* focado em *food delivery*. No Brasil, a maior empresa desse segmento é o *iFood*, com mais de 80% do mercado no país (Josué de Oliveira, 2022).

Houve um crescimento acelerado nos negócios de *food delivery* por conta da pandemia da COVID-19, resultando globalmente em um aumento de 88% nas sessões em aplicativos de delivery em 2020 comparado a 2019, e esse aumento continuou em 2021, para mais 43% (Tiahn Wetzler, 2021). Diversos são os benefícios para um restaurante de estar presente em um *marketplace*, como não precisar se preocupar com questões de local físico como decoração e construção de toda uma parte específica do estabelecimento para que os consumidores possam ir fisicamente ao local. Além disso, existe uma grande exposição ao estar presente em um aplicativo que já é acessado por milhares de pessoas, sem

ter então a necessidade de criar um sistema próprio e construir um público de maneira totalmente individual (Katie McCann, 2020).

Esse aumento em *food delivery* fez com que uma modalidade específica de restaurante apresentasse um aumento muito grande durante esse tempo de pandemia, as chamadas *dark kitchens*, que são estabelecimentos focados unicamente em *delivery*, ou seja, não apresentam um local físico para os clientes consumirem seus produtos (Deliverect, 2023). Esses restaurantes em sua maioria são pequenos e possuem poucas pessoas trabalhando em seu local, facilitando o gerenciamento para o dono do estabelecimento.

Essa possibilidade de abrir um restaurante na própria casa da pessoa e conseguir promover o mesmo para uma ampla variedade de consumidores através de aplicativos de *food delivery* pode, no entanto, se mostrar como uma armadilha para pessoas que não possuem um conhecimento de questões gerenciais. Um planejamento bem feito é o principal ponto para que um restaurante se mantenha em boas condições de funcionamento, como demonstrado em estudos como (PARSA et al., 2011), onde são identificados os principais pontos de falhas em restaurantes destacando questões como a falta de capital financeiro e a incompetência empresarial dos donos do estabelecimento, e em outras análises como visto em (SELF; JONES; BOTIEFF, 2015), onde é pontuado que um *business plan*, ou seja, um plano de negócios bem feito, é um dos principais motivos para que empresas continuem existindo, e tal planejamento passa por um ponto crucial que é o controle do seu estoque de insumos (SISCHEF, 2023).

Como os restaurantes já estão inseridos dentro de uma plataforma em que conseguem controlar seus pedidos, que é o caso dos aplicativos de *food delivery*, surgiu a oportunidade de que esses aplicativos também pudessem oferecer aos seus estabelecimentos cadastrados uma maneira de comprarem os insumos necessários para tal restaurante. Assim é possível realizar a conexão entre restaurantes e consumidores via *delivery*, assim como uma conexão entre restaurantes e fornecedores de insumos, dentro do mesmo sistema. Com o restaurante fornecendo as informações de pratos e detalhes de pedidos para o aplicativo, é vista uma oportunidade clara em que um sistema de recomendação poderá agir usando tais informações. É possível, por exemplo, mostrar para o dono do estabelecimento quais os produtos que ele precisa comprar, baseando-se no cardápio inserido pelo estabelecimento dentro do aplicativo, assim como também nas vendas que estão sendo feitas para os clientes. Além disso, é possível indicar potenciais locais para a compra de tais produtos.

1.1 Justificativa

A área de *CRM* (*Customer relationship management*) de uma empresa é a parte responsável por gerenciar todas as interações que a empresa realiza com os seus clientes,

gerando assim engajamentos para algum produto específico, construindo uma nova base de potenciais clientes ou acelerando as vendas da empresa, seja ofertando melhores produtos ou criando campanhas de *marketing* informando descontos para seus clientes (SALESFORCE, 2023a). Porém, encontrar a melhor base de clientes, assim como entender o que cada cliente específico espera receber da empresa, é um desafio que nem todas as equipes conseguem gerir. De acordo com pesquisa feita pela *SKUid*¹, aproximadamente um terço dos projetos de *CRM* falham (SALESFORCE, 2023b), gerando um desgaste grande nas empresas que tentam implantar tais projetos para melhorarem as comunicações com seus clientes. A criação do sistema de recomendação proposto neste projeto possibilitará uma comunicação mais assertiva da empresa com os seus restaurantes, facilitando assim uma possível conversão desse usuário dentro da plataforma, impactando em KPI's gerenciais de vendas.

Este projeto de TCC será realizado em parceria com a empresa líder em delivery de comidas no Brasil que fornecerá os dados para a realização do mesmo. A justificativa para a realização do mesmo é de alavancar os resultados dessa empresa em questão de vendas e comunicações assertivas com os restaurantes, se tornando uma ferramenta mais atrativa para que eles possam fazer as compras de insumos e bebidas dentro da plataforma fornecida pela empresa.

O banco de dados da empresa possui diversos restaurantes já cadastrados ao redor de todo o Brasil e cada um possui o seu próprio cardápio de bebidas e pratos que são oferecidos para os seus consumidores. Esse cardápio é cadastrado pelo próprio restaurante, com informações de nome do prato e descrição do prato. Diariamente os restaurantes vendem aos seus consumidores esses produtos de forma online, gerando assim uma quantidade valiosa de dados para a empresa, como por exemplo os pratos mais vendidos por cada restaurante específico.

Dentro da empresa também existe o cadastro de fornecedores de alimentos, desde mercados menores a grandes atacados. Esses fornecedores se cadastram na plataforma com o objetivo de oferecer aos restaurantes insumos para o funcionamento dos seus negócios, ofertando bebidas e alimentos de uma forma online semelhante a um *marketplace*.

A relação entre os fornecedores e os restaurantes é realizada por meio de comunicações que são feitas pela equipe de *marketing* da empresa. Tais comunicações, como disparos de mensagens via *Whatsapp* e também email oferecem para o cliente (restaurante) algumas promoções que podem estar acontecendo na plataforma, como algum insumo importante sendo ofertado por algum fornecedor com um preço menor.

Dado esse cenário, surge uma oportunidade muito clara para aprimorar as comunicações feitas para os restaurantes, visto que a empresa possui os dados de cardápio e

¹ <https://www.skuid.com/>

vendas dos mesmos, assim como as informações sobre insumos ofertados pelos fornecedores já cadastrados na plataforma. No entanto, a empresa não possui um sistema automático de recomendação inteligente de insumos para esses restaurantes, avaliando, por exemplo, os seus cardápios, para fornecer uma ligação realmente efetiva entre o que ele consome e o que ele vai comprar.

1.2 Objetivos

O objetivo deste projeto é a criação de um sistema de recomendação de melhores fornecedores para restaurantes cadastrados no aplicativo de *delivery iFood* a partir de quantos insumos e ingredientes os fornecedores e restaurantes possuem em comum.

A partir desse catálogo será aplicada técnicas de mineração de texto para retirar os possíveis itens que estão presentes dentro de um prato dado o seu nome e descrição fornecidos pelos restaurantes. Com essas informações extraídas será recomendado para os restaurantes um conjunto de produtos que os fornecedores cadastrados na plataforma possuem que sejam mais condizente com o que o restaurante realmente precisa, baseado nesses itens presentes dentro do seu catálogo de ofertas.

1.2.1 Objetivos Específicos

- Extração de duas bases de dados, sendo uma de restaurantes e seus pratos vendidos no *iFood* e outra base com os fornecedores e seus insumos vendidos no *iFood Shop*
- Pré-processamento textual aplicado nas duas bases, utilizando *tokenização* e remoção de palavras desnecessárias para o projeto
- Transformação de pratos em ingredientes transformando o cardápio dos restaurantes no máximo possível de ingredientes específicos que possam ser utilizados pelos pratos
- Agrupamento de restaurantes com os seus ingredientes e de fornecedores com seus insumos
- Realizar a recomendação de fornecedores para restaurantes baseado no quanto os dois possuem ingredientes e insumos em comum

1.3 Organização do Trabalho

Este trabalho está sendo organizado em capítulos, sendo:

- **Capítulo 2 - Revisão Bibliográfica:** é apresentada toda a fundamentação com explicações do porque este trabalho foi desenvolvido, demonstrando a história das

vendas feitas de forma *online* até chegar no universo de *marketplaces* que temos hoje em dia. Também são explicadas as diversas dores e oportunidades que os restaurantes possuem nos dias atuais, onde temos uma influência muito grande por parte dos aplicativos de *delivery* dentro desses estabelecimentos

- **Capítulo 3 - Desenvolvimento** - todas as etapas técnicas que foram aplicadas neste trabalho são explicadas neste capítulo, demonstrando os diversos pré processamentos que foram necessários para a limpeza da base de dados
- **Capítulo 4 - Resultados** - aqui é demonstrado como os pré processamentos explicados no capítulo 3 afetaram a base de dados trabalhada, transformando os textos apresentados dentro do aplicativo de *delivery* em ingredientes e, posteriormente, recomendações de fornecedores e seus insumos.
- **Capítulo 5 - Conclusão** - o último capítulo deste trabalho apresenta todos os resultados e conclusões obtidas com a finalização do projeto

2 Revisão Bibliográfica

Este Capítulo apresenta todo o contexto necessário para entender o motivo de construção e elaboração deste estudo, mostrando também, no final, os trabalhos que estão relacionados com o tema proposto.

As Seções 2.1 e 2.2 apresentam como surgiu todo o contexto de estabelecimentos vendendo produtos no mundo *online* e o início das primeiras vendas feitas de forma não física. A seção 2.3 e 2.4 mostram o surgimento dos aplicativos de entrega e como a pandemia da COVID-19 mudou completamente essa área de *delivery*.

Na Seção 2.5 são apontados os principais problemas que um gestor pode ter na gerência de estabelecimentos. Tais problemas podem ser amenizados utilizando técnicas de mineração de texto que são explicadas nas seções 2.6, 2.7 e 2.8.

Por fim, a seção 2.9 apresenta os trabalhos relacionados que possuem também técnicas de mineração de texto e sistemas de recomendação focados para restaurantes.

2.1 Lojas Virtuais

Com o avanço da tecnologia tornou-se possível que empresas pudessem se conectar com os seus consumidores de forma *online*. Esse avanço afetou diretamente os estabelecimentos, que tinham como seu principal objetivo a venda de algo para as pessoas, que antes acontecia apenas de forma física. Thomai Serdari, professor na *New York University's Stern School of Business*, demonstrou que os anos 90 foram de extrema importância para que o comércio *online* fosse viável e lucrativo para as empresas (LUFKIN, 2020).

Vendo que o mundo estava cada vez mais conectado à Internet, as empresas começaram a entrar em uma estratégia de construção de lojas virtuais (do inglês, *online store*), ou seja, criar os seus próprios sites e plataformas para mostrarem seus produtos aos clientes e proporcionar que os mesmos possam realizar compras de forma *online*, sem a necessidade de irem em lojas físicas. É dito que essa modalidade de compras não presenciais começou em 1979, quando Michael Aldrich, um inventor inglês, criou um sistema em que conectava a televisão de casa com o computador de um estabelecimento via telefone domiciliar para fazer compras sem a necessidade do consumidor se dirigir ao estabelecimento (NORMAN, 2023).

Em 1994 uma grande rede de pizzas dos Estados Unidos, *Pizza Hut*, iniciou uma revolução utilizando os conceitos propostos por Michael Aldrich (LUFKIN, 2020). Nesse ano, a *Pizza Hut* mostrou para seus clientes, pela primeira vez, a oportunidade de realizar compras via *PizzaNet*, um portal criado pela empresa em que o cliente poderia realizar o

seu pedido sem precisar se dirigir especificamente a algum restaurante da rede.

Apesar da facilidade que as lojas virtuais geraram para os consumidores, nem tudo é vantajoso nessa estratégia olhando para os negócios em si. Um dos grandes problemas na criação de uma plataforma online para um negócio próprio é o custo para que isso seja viável para um pequeno ou médio empreendimento (PORTER, 2023). É necessário um trabalho para criar a plataforma, personalizar ela da forma que represente bem o negócio, além de que a plataforma precisa chegar ao consumidor final, gerando assim um trabalho grande de *marketing* para que se atinja o consumidor correto para o negócio (PARMAR, 2021). Todo esse trabalho para a criação de uma loja virtual própria gera um empecilho para empresas que não possuam disponibilidade tanto financeira como de equipes, já que é necessário pessoas com conhecimentos específicos de tecnologia para a construção da plataforma *online*. A principal forma de se apresentar ao mundo do *ecommerce* sem se preocupar com todo o trabalho de criação própria de uma plataforma é a entrada do negócio dentro de um *marketplace*.

2.2 Marketplace

Uma das principais formas de levar mais visibilidade para um comércio é através de grandes *marketplaces*, ou seja, plataformas onde diversos vendedores são reunidos para vender seus produtos para os clientes via internet (VIDAL, 2019). A *Amazon*¹, maior empresa de *marketplace* do mundo, divulgou em seus resultados anuais que 60% de suas vendas são realizadas por vendedores independentes (PYMNTS, 2023), ou seja, pequenas e médias empresas que se colocam em uma grande visibilidade ao se cadastrarem nessas plataformas de *marketplace*.

É comum pensar em *marketplaces* e visualizar apenas as vendas de produtos domiciliares ou de livros, porém os produtos oferecidos tem mudando conforme o acesso a internet foi se popularizando. Nos dias atuais temos *marketplaces* de diversos estilos. Por exemplo, o *Airbnb*² é uma das principais empresas de *marketplace* do mundo em que os usuários conseguem encontrar diversas hospedagens ao redor do mundo, sem que seja necessária a pesquisa, por exemplo, de cada hotel específico para uma viagem. Já a *SHEIN*³ foi considerada a marca de moda mais popular em todo o mundo em 2022 (HAQQI, 2022).

2.3 Food Delivery

O conceito de *marketplace* então se expandiu para a área de alimentação. Como apresentado na Seção 2.1, a *Pizza Hut* foi precursora no que se diz respeito a lojas virtu-

¹ <https://www.amazon.com.br/>

² <https://www.airbnb.com.br/>

³ <https://pt.shein.com/>

ais, mas na época ainda era uma plataforma específica da rede de pizzas estadunidense. O primeiro serviço que agregava diversos restaurantes em uma única plataforma surgiu em 1995, chamado *World Wide Waiter*⁴ (JACKSON, 2021). Na época o jornal *Palo Alto Online* mostrou o quão original e inovador era o *Waiter*, possibilitando que os consumidores visualizassem os menus de 60 restaurantes diferentes e até mesmo separando eles por localizações e tipos de cozinha específicos (HARRIS, 1996).

O mercado de *food delivery* é um dos mais competitivos dos diferentes tipos de *marketplaces* existentes. Diversas são as empresas que disputam a atenção do consumidor quando se trata dessa modalidade. Porém nem todas as plataformas conseguem se manter líderes ao redor de todo o mundo, como é o caso da dominância global da *Amazon* no que se refere a *marketplace* de livros, por exemplo. Uma prova disso é o relatório providenciado pela (DATA.AI, 2021) na parte de *delivery* de alimentos. Pode-se observar por ele que a dominância nesse segmento em cada país é feita por diferentes empresas. Na Argentina é visto como líder o aplicativo *PedidosYa*⁵, nos Estados Unidos o *DoorDash*⁶, no México o *UberEats*⁷ e no Brasil o *iFood*⁸.

A concorrência no ramo de alimentação é muito forte, visto as diferenças de preferências entre os países. Gigantescas empresas globais, como é a *Uber*, não conseguem manter o seu ramo de *food delivery* em todos os países, como foi o caso da paralização de seus serviços de entrega de comida no Brasil (G1, 2022), mesmo tendo uma participação de 13% no serviços de *delivery* no país (GRANDI, 2022). No Brasil, o *iFood* é o maior aplicativo de *food delivery* do país e é considerado o sexto maior do mundo (IFOOD, 2022), chegando a ser responsável por mais que 80% das vendas de *delivery* que acontecem no país (ROSA, 2022).

Diversos são os benefícios para o consumidor final no que se diz respeito a *delivery* de comida. Essa inovação leva muita praticidade para o consumidor, retirando a necessidade de realizar todo o processo de preparação de uma refeição. Além disso, essa praticidade de receber um produto sem precisar se locomover acontece mesmo pagando um pouco mais caro, como mostra a pesquisa da (ABRASEL, 2022), em que reportaram um aumento de 17.5% no preço das refeições dentro da plataforma do *iFood*, comparado ao preço das mesmas refeições no restaurante presencial.

Já para os restaurantes existem também diversos benefícios em utilizarem uma plataforma de *food service*, sendo eles:

- Aumento na receita: no artigo publicado por (KIMES; LAQUE, 2011), os donos dos

⁴ <https://www.waiter.com/>

⁵ <https://www.pedidosya.com/>

⁶ <https://www.doordash.com/>

⁷ <https://www.ubereats.com/br>

⁸ <https://www.ifood.com.br/>

restaurantes procuravam serviços de *food delivery* principalmente para o aumento nas receitas. Isso se dá através de alguns pontos, tais como, o aumento no preço dos pratos oferecidos de forma online, quando comparados aos preços trabalhados de forma física, o aumento no volume de pedidos, já que o restaurante ficaria bem mais disponível para os consumidores através de forma online, sem a necessidade de irem a um local físico, e por fim o aumento na frequência dos pedidos, já que consumidores satisfeitos com algum pedido se tornam mais frequentes a repetirem os mesmos pedidos;

- Localização: outro principal benefício de estar presente em uma plataforma de pedidos online é a questão da localização do restaurante não importar tanto para o consumidor final, como é mostrado em (SEE-KWONG et al., 2017). Em grandes centros de cidades existem diversos restaurantes disponíveis para os consumidores. Porém, um grande ponto negativo de estar presente em áreas mais tumultuadas é a aglomeração de várias pessoas ao mesmo tempo, o que afeta diretamente questões como segurança e trânsito;
- Exposição ao consumidor: também em (SEE-KWONG et al., 2017), é visto que a exposição por estar em uma grande plataforma online de *delivery* é muito benéfica para os restaurantes. Caso um estabelecimento não esteja presente nessas plataformas ele precisará, nos dias atuais, promover o seu restaurante de forma própria e bem mais extensiva, para concorrer com outros estabelecimentos que possam estar dentro dos grandes *players* de *food delivery*

2.4 Pandemia e *Dark Kitchens*

Uma das principais áreas impactadas pela pandemia da COVID-19 foram os restaurantes, como mostra uma pesquisa da (ABRASEL, 2021a) em que 40% dos restaurantes especializados em comida a quilo fecharam no Brasil, gerando uma perda de aproximadamente 80 mil estabelecimentos. Em outra pesquisa, feita pela ANR (2020), 22% das empresas do setor de restaurantes não iriam resistir à pandemia, além de que 64% do mercado de restaurante já haviam promovido demissões por conta da dificuldade de se manterem abertos e com clientes.

Com a proibição de aglomeração praticada durante a pandemia, o chamado *lock-down*, os restaurantes ficaram impossibilitados de receberem os clientes em seus locais físicos. Tendo uma falta de perspectiva otimista para a volta dos consumidores, muitos aderiram aos serviços de entrega de alimento, o chamado *delivery*. Como apontado em (ABRASEL, 2021b), 89% dos restaurantes brasileiros adotaram a modalidade de *delivery* por conta da pandemia da COVID-19. Em uma pesquisa de 2021, a ANR (2021) verificou com diversos responsáveis por estabelecimentos de bares e restaurantes e notou que

até 40% do faturamento das empresas na época vinha da modalidade de *delivery*. Isso representa um aumento de 16 pontos percentuais em comparação com o que era antes da pandemia. Além disso, a pesquisa revelou que 85% dos restaurantes irão continuar investindo na entrega a domicílio para os clientes, não ficando apenas nas suas operações presenciais.

Com o visível aumento de *food delivery* durante a pandemia, surgiu também o conceito de *dark kitchen*, ou "restaurante fantasma". Esses estabelecimentos são chamados assim por não apresentarem um local específico para consumo dos clientes, sendo um tipo de negócio que funciona apenas com o objetivo de *delivery* para o consumidor (SCUADRA, 2020). Basicamente, os restaurantes fantasma são apenas uma cozinha central, sem a necessidade de existir um salão ou de precisar de contratação de garçons, como é a prática normal de um restaurante com local físico para atendimento ao cliente. E o número de restaurantes que funcionam dessa forma não é pequeno, como mostra o estudo feito por Hakim et al. (2023), em que até 35% dos restaurantes analisados em São Paulo são considerados *dark kitchen*, e muitas vezes os consumidores nem mesmo sabem que estão pedindo de estabelecimentos que funcionam dessa maneira.

A combinação da pandemia da COVID-19, *food delivery* e *dark kitchens* abriu portas para que pessoas que não possuam muita condição financeira pudessem abrir um restaurante menor dentro de suas próprias cozinhas. Com os aplicativos de entrega, elas conseguem mostrar o estabelecimento para diversos consumidores, que já estejam cadastrados na plataforma, e não precisam mais se preocupar em construir um local específico para o restaurante. O tempo e os custos que levavam a criação de estratégias de *marketing* específicas para atrair consumidores são reduzidos, já que o restaurante consegue se apresentar para muitos clientes através do cadastro dentro das plataformas de *delivery*. Um exemplo disso é a estratégia adotada por uma rede de comida brasileira por quilo no Brasil, a Divino Fogão. Para abrir uma franquia tradicional da empresa era necessário um investimento de R\$700 mil reais, no mínimo, 35 metros quadrados e 18 funcionários, enquanto para abrir uma "*versão dark kitchen*", uma pessoa precisaria desembolsar apenas R\$8 mil reais e nem mesmo existe a exigência de um número específico de funcionários ou de tamanho (AMERICO, 2021).

2.5 Dificuldades de restaurantes

Com o aumento das facilidades para a criação de restaurantes, com a combinação de *dark kitchens* e *food delivery*, existe uma preocupação em relação às dificuldades que um restaurante pode ter. Em uma pesquisa feita pela Associação Brasileira de Bares e Restaurantes (Abrasel), cerca de 35% dos novos empreendimentos nessa área fecham antes mesmo do primeiro ano de abertura, e esse número sobe para 50%, quando se analisa os

dois primeiros anos (MAGAZINE, 2015).

Diversos são os problemas que levam ao fechamento precoce de estabelecimentos especializados em comida. Em "*Why Do Restaurants Fail? An Analysis of Macro and Micro Factors*", o autor (PARSA et al., 2011) comenta alguns dos principais motivos que causam esse grande número de falhas em empreendimentos de restaurantes. Na parte de fatores macro, é citado problemas como questões econômicas do país, já que o consumo de alimentos fora de casa fica mais caro do que fazer, por exemplo, marmitas em casa (OTEMPO, 2022). Além disso, há as questões de legislação, que podem alterar a forma como um restaurante deve ou não funcionar, afetando por exemplo, questões tributárias de um estabelecimento.

Já considerando os fatores micro que afetam diretamente as falhas nesses empreendimentos, existe a questão da localização de um restaurante, que dependendo do local físico, pode não ser atrativo para os clientes. Além disso, existem as questões complexas relativas a criação de uma marca forte para o restaurante, e todo o processo de *marketing* necessário para dar visibilidade. O principal problema de vários restaurantes não conseguirem se manter é o fator do capital financeiro de quem vai iniciar esse tipo de empreendimento. É fundamental que um restaurante tenha as questões financeiras muito bem controladas, e isso está relacionado principalmente com o controle de estoque (PARSA et al., 2005).

Muitos donos de restaurantes ainda realizam suas compras de forma física, verificando os fornecedores de sua região e fazendo manualmente as cotações dos preços dos produtos (RUPP, 2020). Com o objetivo de levar mais benefícios para o restaurante cadastrado em sua plataforma, e facilitar essa busca por preços de vários fornecedores, o *iFood* iniciou um projeto chamado *iFood Shop*⁹. Este projeto tem o objetivo de construir uma plataforma de conexão entre fornecedores de alimentos e os estabelecimentos que estejam cadastrados em sua base para ajudar no controle de estoque dos restaurantes, proporcionar um preço menor do que o que é ofertado em lojas físicas e retirar a tarefa do comerciante precisar averiguar em diferentes lojas os preços de seus produtos. Assim, o plano da empresa é a construção de um *marketplace* de fornecedores para os restaurantes cadastrados em sua plataforma, criando uma cadeia completa para o dono de um estabelecimento. Dessa forma, os restaurantes vão conseguir vender seus pratos e também comprar seus insumos através de uma plataforma única.

2.6 Mineração de Texto

Para a construção de uma plataforma de recomendação de fornecedores de insumos para restaurantes é necessária a aplicação de técnicas de mineração de texto a partir dos

⁹ <https://shop.ifood.com.br/>

cardápios que são cadastrados pelos estabelecimentos dentro das plataformas. São esses cardápios, apresentados em forma de texto, que fornecerão os dados necessários para a construção dessa recomendação. A mineração de texto (em inglês, *Text Mining*) é uma área da ciência da computação com foco na extração de informações que possam ser úteis para um determinado problema mas que estão apresentadas em dados textuais não estruturados (STEDMAN, 2023) como, por exemplo, os cardápios dos restaurantes cadastrados em aplicativos de entrega.

Dados não estruturados não possuem uma visualização amigável como dados estruturados, que ficam apresentados em linhas e colunas em um formato de tabela, em que cada linha apresenta um tipo de informação específica e cada coluna possui também um tipo de dado específico para ser apresentado (SMALLCOMBE, 2023). Apesar da facilidade de trabalho com dados estruturados, cerca de 80% dos dados mundiais serão no formato de dados não estruturados em até 2025 (DIALANI, 2020).

Muitas são as áreas em que se pode utilizar a aplicação de mineração de texto. No âmbito de segurança online, são aplicadas técnicas nas mensagens trocadas por usuários para a detecção prévia se o conteúdo de tal mensagem pode ser nocivo ou não, que seria a detecção, por exemplo, de um *spam* em uma caixa de e-mail (EXPERT.AI, 2022). Pode-se usar mineração de texto até mesmo para detectar o chamado cibercrime, ou seja, crimes que são cometidos de forma *online*, como foi mostrado recentemente no Brasil em que diversas contas na rede social *Twitter* foram apagadas por apresentarem textos que poderiam estar amplificando ataques terroristas em escolas (AMARAL, 2023).

2.7 Etapas da Extração de Conhecimento a partir de Textos

Extração de conhecimento (em inglês, KDD - *Knowledge Discovery in Databases*) é todo o processo de coleta, transformação e refinamento que é aplicado em um conjunto de dados até que esses virem informações úteis para algum objetivo (SHARMA, 2020). O KDD é composto de um conjunto de etapas, que serão descritas a seguir. Será dado enfoque ao processo de extração de conhecimento a partir de texto.

2.7.1 Coleta de dados

A primeira etapa do processo de KDD é a coleta de dados (do inglês *Data Collection*), onde é identificado qual o conjunto de dados que será processado e analisado pelas próximas etapas (CHEHAB, 2020). Para essa primeira etapa, é necessário ter em mente qual será o problema analisado e quais as expectativas de solução do mesmo, para que se faça uma boa extração inicial dos dados. Normalmente, uma base inicial irá conter informações que podem não ser necessárias para o resto de um projeto. Assim, diminuir essa base é crucial para que as próximas etapas não apresentem problemas de desempenho

ou que dados desnecessários afetem os resultados de avaliações que serão feitas tendo a presença de *outliers*, ou seja, dados que se dispersam muito da realidade média da base selecionada (LEMONAKI, 2021).

2.7.2 Pré-Processamento

Após a coleta inicial dos dados, e feita uma filtragem a partir deles para que se apresente apenas informações que possam ser importantes para uma análise, é então realizada a etapa de pré-processamento. Esta é uma etapa que irá preparar o dado para a etapa de mineração, que virá em seguida. Isso se deve ao fato de que, na maioria das vezes, os textos trabalhados não irão apresentar de forma direta as informações que são necessárias para a aplicação de técnicas de mineração em cima deles. Assim como nem todo o conteúdo presente em um texto será útil para a análise.

Uma das técnicas de pré-processamento focado em texto é a detecção e remoção das palavras chamadas de *stop words*, ou "palavras de parada". Essas palavras não apresentam função semântica nos textos, e são classificadas como artigos, preposições e pronomes (PORTUGUÊS, 2022). Uma das formas mais clássicas de se retirar essas *stop words* de um texto é utilizar uma lista já montada contendo todas as possíveis palavras que não representem uma importância para a análise de um texto. A parte do pré-processamento então é realizada para detectar e remover anomalias além de capturar a real essência de um texto (VIJAYARANI et al., 2015).

Outra etapa que faz parte do conjunto de processos dentro de um pré-processamento é a chamada *stemming*. O principal propósito dessa etapa é reduzir as várias formas gramaticais diferentes que uma palavra pode possuir, como verbos, advérbios, adjetivos, para apenas ler a sua forma raiz (JIVANI et al., 2011). Essa parte do pré-processamento serve também para reduzir a quantidade de palavras que serão analisadas por algum sistema, já que todas as suas formas gramaticais, apesar de apresentarem escritas levemente diferentes, partem sempre do início da palavra.

2.7.3 Mineração de Dados

Mineração de dados, em inglês, *data mining*, é uma combinação de análises estatísticas, aprendizado de máquina e gerenciamento de dados para extrair informações úteis de uma base de dados grande (THURASINGHAM, 2000). As aplicações de técnicas de mineração de dados podem ser usadas em diversas áreas como medicina, finanças, controles de finanças e também em questões de logística. No projeto que está sendo apresentado para o *iFood*, a Mineração de Dados pode ser útil para realizar classificações de ingredientes dos pratos dos restaurantes, agrupamento de pratos similares assim como criar regras de associação para que se possa relacionar dois ingredientes com nomes diferentes mas

que sejam do mesmo estilo.

Existem diversos algoritmos e técnicas que são utilizados na de mineração de dados. Uma dessas técnicas é a classificação que tem como objetivo alocar um item em algum conjunto pré definido de classes ou grupos (KESAVARAJ; SUKUMARAN, 2013). Um dos exemplos clássicos de utilização de classificação é verificar o risco de liberação de crédito para um cliente específico dado seu histórico comportamental de movimentações financeiras. Dado um conjunto pré determinado de usuários já classificados pode-se construir um modelo de classificação para, automaticamente, alocar esse novo cliente no seu respectivo nível de risco de fraude quando comparado a clientes similares.

Para se determinar o quanto dois dados estão relacionados ou não são utilizadas as chamadas regras de associação, muito utilizada em questões relacionadas a comércio para se relacionar produtos úteis para os consumidores (CÔRTEZ; PORCARO; LIFSCHITZ, 2002). Para que uma regra de associação, que existe na forma $X \rightarrow Y$ (lê-se: Se X então Y), seja considerada interessante para um problema, ela precisa atingir um nível mínimo de *suporte* e *confiança*, duas métricas que são utilizadas para que se possa afirmar que realmente um dado induz a outro (ZHANG; WU, 2011).

Empresas como os aplicativos de *food delivery* possuem em seus bancos diversos dados acerca de seus usuários, e a transformação desses dados em informações é capaz de gerar tomadas de decisões mais estratégicas para a empresa como um todo, e a mineração de dados existe, com uma série de técnicas, para gerar tais tomadas de decisões mais corretamente o possível. Nos dias atuais as empresas possuem equipes de *CRM* (*Customer Relationship Management*, ou gestão de relacionamento com o cliente) que utilizam das informações geradas por um trabalho realizado a partir de mineração de dados para realizar análises de mercado, segmentações dos clientes, customização de produtos para bases específicas, análise de fraudes, entre outros (WILSON, 2022).

2.8 Sistemas de recomendação

É denominado sistema de recomendação qualquer sistema que se baseie em dados para analisar gostos e preferências de seus usuários para, assim, realizar outras recomendações para manter o usuário dentro do sistema (BETANCOURT; ILARRI, 2020). Definidos os gostos de um usuário, o principal objetivo de um sistema de recomendação é mostrar para ele produtos similares ao que ele já demonstrou interesse (HOLEWA, 2020). A principal qualidade em uma recomendação é o quão "relacionado" é uma operação já feita anteriormente por um usuário com a operação que o sistema está predizendo. O desafio então é descobrir padrões que possam ser benéficos para os usuários do sistema (SMITH; LINDEN, 2017).

Dentre as principais redes sociais e *marketplaces* existentes se faz presente o uso

de sistemas de recomendação. Em 2010, o *Youtube*¹⁰ iniciou o uso de sistemas de recomendação para proporcionar aos seus usuários vídeos que poderiam ser de seus interesses, baseado em seus históricos de vídeos, previamente assistidos. A adoção dessa estratégia foi muito benéfica para a empresa, gerando um resultado surpreendente em que 60% dos vídeos clicados em sua página inicial eram vídeos de recomendação para os usuários (DAVIDSON et al., 2010).

A criação de um sistema de recomendação de fornecedores para restaurantes ajuda em uma das mais vitais etapas de gerenciamento de um estabelecimento, que é a escolha correta dos fornecedores e de seus produtos, visto que os restaurantes procuram sempre: preços competitivos, boa qualidade na entrega do pedido e cooperação entre o fornecedor e o estabelecimento (FLOREZ-LOPEZ, 2007). Porém, não são encontrados trabalhos similares que tentam ajudar o estabelecimento em si. Muitos trabalhos de recomendação, que possuem os restaurantes como uma das peças relacionadas em sua estrutura, baseiam-se na recomendação dos restaurantes para os consumidores, tendo então o consumidor final como o principal foco da recomendação.

2.9 Trabalhos relacionados

Existem diversos trabalhos que buscam extrair conhecimento a partir de bases de dados de texto. São diversos algoritmos e medidas de avaliação que podem ser usados a fim de avaliar a qualidade dos resultados produzidos em cada uma das etapas da extração do conhecimento. Os trabalhos relacionados são apresentados na Tabela 1.

O trabalho proposto por Kurnia et al. (2019) demonstra como os dados das vendas de restaurantes podem ser usados para se obter benefícios. Esse estudo, feito em cima de uma base de dados com informações transacionais privadas de vendas de um restaurante chamado *O! Fish*, permitiu, via análise de dados, promover um novo tipo de *combo* de produtos dentro do restaurante com o propósito de garantir um *upsell*¹¹ nas vendas do estabelecimento. O artigo utiliza conceitos de ciência de dados como *regras de associação*, e medidas como suporte e confiança, além de utilizar o algoritmo *Apriori* sendo a forma mais apropriada para trabalhar com o nível de informações que ele possuía. Para o cálculo e a amostragem das medidas como suporte, o autor utilizou a ferramenta *Rapid Miner*, que é usada para analisar dados e encontrar associação entre as informações fornecidas. O estudo foi bem sucedido, encontrando novos *combos* de produtos vendidos no restaurante, possibilitando assim novas opções para os clientes.

A partir de uma base de dados de *reviews* feitas por consumidores a respeito de restaurantes extraída do site *Tripadvisor*, Asani, Vahdat-Nejad e Sadri (2021) mostram

¹⁰ <https://www.youtube.com/>

¹¹ Técnica para aumentar a lucratividade em cima da venda para o cliente

Tabela 1 – Tabela com trabalhos relacionados

Trabalho	Pré-processamento	Algoritmo	Medidas de Avaliação	Base de Dados
(KURNIA et al., 2019)	Remoção de Stop Words	Apriori	Suporte e Confiança	Base privada de um restaurante
(ASANI; VAHDAT-NEJAD; SADRI, 2021)	Tokenização; Stemming; Clusterização	Wu Palmer	Jaccard; F-Measure; Precisão	Base de avaliações do TripAdvisor
(LIMA; GONÇALVES, 2022)	-	Algoritmo Semântico	Levenshtein; Jaro-Winkler; Jaccard	Base pública do IBGE
(MEIRELLES; GONÇALVES; GOMES, 2021)	Remoção de Stop Words; Stemming	Algoritmo Semântico	Acurácia; Jaccard; Levenshtein	Base pública do IBGE
(RAKHMAWATI; JANNAH, 2021)	Tokenização; Stop Words	Algoritmo Semântico	Jaccard; Jaro Winkler; Levenshtein	Base do Open Food Facts

como comentários feitos na internet podem ser úteis para influenciar positivamente as vendas de restaurantes. Esse trabalho objetivou analisar como os clientes se sentiam sobre alguns alimentos baseado em seus comentários feitos no site *Tripadvisor*¹². Para a construção do projeto, os autores precisaram então utilizar técnicas de mineração de texto para a extração das palavras e classificar se o comentário feito sobre um restaurante era de tendência positiva ou negativa, além da necessidade de agrupar certos alimentos para que eles ficassem em uma hierarquia única. Para isso o trabalho precisou organizar os alimentos em conjuntos que possam ser similares, unindo em um mesmo grupo itens como "camarão" e "ostra", enquanto alimentos como "frango, filé" ficariam em outros grupos. O projeto então utiliza os comentários publicados no site de avaliação e verifica se as frases dos usuários, acompanhadas pelos alimentos, são positivas ou negativas com base em técnicas de mineração de dados chamada de análise de sentimentos. Com essas relações sendo feitas, como, por exemplo, determinar que o usuário escreveu uma avaliação positiva ao mesmo momento em que ele estava se referindo a um prato do restaurante que é composto por frango, o projeto conseguiu recomendar outros restaurantes com os mesmos grupos analisados pelo trabalho em que o cliente tenha tido uma reação positiva. Nesse exemplo, o usuário seria recomendado a outros restaurantes que também possuam pratos relacionados a frango.

No trabalho de Lima e Gonçalves (2022), os autores propõem calcular a similaridade semântica entre nomes de produtos alimentícios para realizar um agrupamento

¹² <https://www.tripadvisor.com.br/>

desses produtos. Como exemplo citado no projeto, itens alimentícios como *batata-baroa*, *mandioquinha* e *batata-inglesa* poderiam ser agrupados em um único conjunto denominado *batata*. No trabalho, foram utilizadas duas *wordnets* do português para comparar o desempenho entre elas. *Wordnet* é um conceito criado dentro da Universidade de Princeton (PRINCETON, 2005) sendo uma base das palavras da língua inglesa e suas relações semânticas entre outras palavras, como sinônimos. Sendo assim, o autor utilizou de uma base de *wordnet* chamada *Open WordNet-PT*, que está disponível na biblioteca *NLTK* da linguagem *Python* e uma outra *wordnet* que foi construída utilizando dois arquivos públicos do IBGE (Instituto Brasileiro de Geografia e Estatística). Foram usadas métricas como *Levenshtein*, *Jaro-Winkler* e *Jaccard* para calcular o quão distante um alimento está de outro a partir de suas similaridades semânticas. Concluiu-se no final de que a *wordnet* gerada a partir dos arquivos do IBGE foram promissoras, em algumas métricas melhores do que utilizando a biblioteca pronta da linguagem *Python*.

O trabalho proposto por Meirelles, Gonçalves e Gomes (2021) propôs também uma análise de pareamento de nomes de produtos e serviços também utilizando medidas de similaridade semântica. Neste trabalho também foram utilizados índices como *Jaccard*, *Levenshtein* e *Jaro-Winkler* para analisar a similaridade entre duas palavras que foram extraídas da base de dados pública do IBGE que é constituída por quase 5.000 descrições de produtos e serviços. No trabalho, os autores não propõem apenas o agrupamento de produtos alimentícios, mas também o agrupamento de despesas de serviços usuais que o Instituto encontra na sociedade brasileira. Para esse trabalho também foram utilizadas técnicas de pré-processamento. Neste trabalho o autor concluiu que a combinação de três medidas de similaridade, como *Levenshtein*, *Jaro* e *Jaccard*, aumentou a acurácia de pareamento. Sendo assim, a combinação dessas três medidas resultou em um melhor resultado do que analisando elas separadamente.

Relacionado a agrupamento de dados referente a alimentos, o trabalho (RAKHMAWATI; JANNAH, 2021) propõe a ideia de agrupar alimentos que possam ser semelhantes a partir da composição presente nos rótulos de ingredientes desses alimentos. O autor utilizou uma base de dados extraída do site *Open Food Facts*¹³, que contém informações sobre um produto e o seu rótulo de ingredientes. Para o trabalho, o autor precisou realizar diversas tarefas de pré-processamento antes de efetivamente agrupar os alimentos similares, ou seja, os dados de ingredientes de um alimento específico foram tratados removendo sinais de pontuação, *stopwords*, realizando uma tradução de ingredientes que, nesse caso, não estavam em língua inglesa. Além disso o autor removeu todos os produtos que estavam sendo duplicados na base do *Open Food Facts* e também removeu produtos que não possuíam os ingredientes cadastrados na base. No projeto apresentado, o autor aferiu que o melhor desempenho para o cálculo de similaridade entre os ingredientes foi

¹³ <https://world.openfoodfacts.org/>

a métrica de distância de *Jaro Winkler*. Percebeu-se também neste projeto, que escritas erradas, como exemplo citado no texto a palavra *pinrapple*, que deveria representar *pineapple*, abacaxi em português, dificultava a mensuração correta de similaridade entre as palavras.

2.10 Considerações Finais

Neste capítulo foi apresentada toda a evolução da *internet* que resultou no surgimento de grandes *marketplaces* que funcionam com robustos sistemas de recomendação criados para aumentarem seus lucros utilizando dados de preferências dos usuários.

Também foram explicados todos os problemas que donos de restaurantes possuem dentro de seus estabelecimentos, com ênfase no quão importante é para um restaurante manter um controle de estoque adequado buscando melhores oportunidades de preço e quantidades de insumos.

Os trabalhos relacionados demonstram que existem diversos projetos com o objetivo de ajudar o restaurante a entender os seus clientes, assim como trabalhos com a proposta de mostrar para os clientes os melhores restaurantes baseado em suas preferências.

No próximo capítulo serão detalhadas as etapas realizadas de mineração de texto para a construção de um sistema de recomendação voltado para os restaurantes para que eles possam construir o estoque de seus suprimentos da melhor forma possível, indicando para o estabelecimento os melhores fornecedores baseado nos cardápios cadastrados no *iFood*.

3 Desenvolvimento

Este capítulo apresenta toda a parte técnica que foi realizada para este projeto, desde a coleta dos dados até a parte de criação de um sistema de recomendação.

A seção 3.1 apresenta um resumo de todas as etapas que compõe este trabalho. Na 3.2 é explicada a coleta inicial dos dados assim como os primeiros pré-processamentos realizados.

Dentro de 3.3 é detalhado um trabalho mais específico de pré-processamento realizado nas bases utilizadas no projeto. Em 3.4 é detalhada como é feita a transformação dos cardápios dos restaurantes em ingredientes específicos.

Em 3.5 é demonstrado o agrupamento final para obter os ingredientes e insumos de cada restaurante e fornecedor específicos.

A seção 3.6 explica como foi realizado o sistema de recomendação e a métrica utilizada para o cálculo de eficiência proposto neste projeto, com a tentativa de unir o melhor fornecedor possível para os restaurantes cadastrados na plataforma de *delivery*.

Por fim a seção 3.7 apresenta um resumo com o objetivo e a ideia por trás de todo o trabalho de mineração de dados aplicado neste projeto.

3.1 Visão Geral da Proposta

Este trabalho possui como objetivo criar um sistema de recomendação de fornecedores para restaurantes, ambos cadastrados na plataforma *iFood Shop*, baseado nos ingredientes que os restaurantes utilizam para a preparação de seus pratos.

O trabalho é dividido nas etapas apresentadas na figura 1 que são detalhadas nas próximas seções. O trabalho se inicia com a coleta da base dos restaurantes e dos fornecedores e, após aplicada toda a etapa de pré-processamento, são realizadas as recomendações dos melhores fornecedores para cada restaurante.

Essa relação criada entre restaurante e fornecedor permite disparar comunicações mais assertivas pela equipe de *CRM* da empresa, deixando claro para os restaurantes quais são os melhores fornecedores para eles, dada essa relação entre ingredientes dos cardápios dos estabelecimentos com os insumos oferecidos pelos fornecedores.

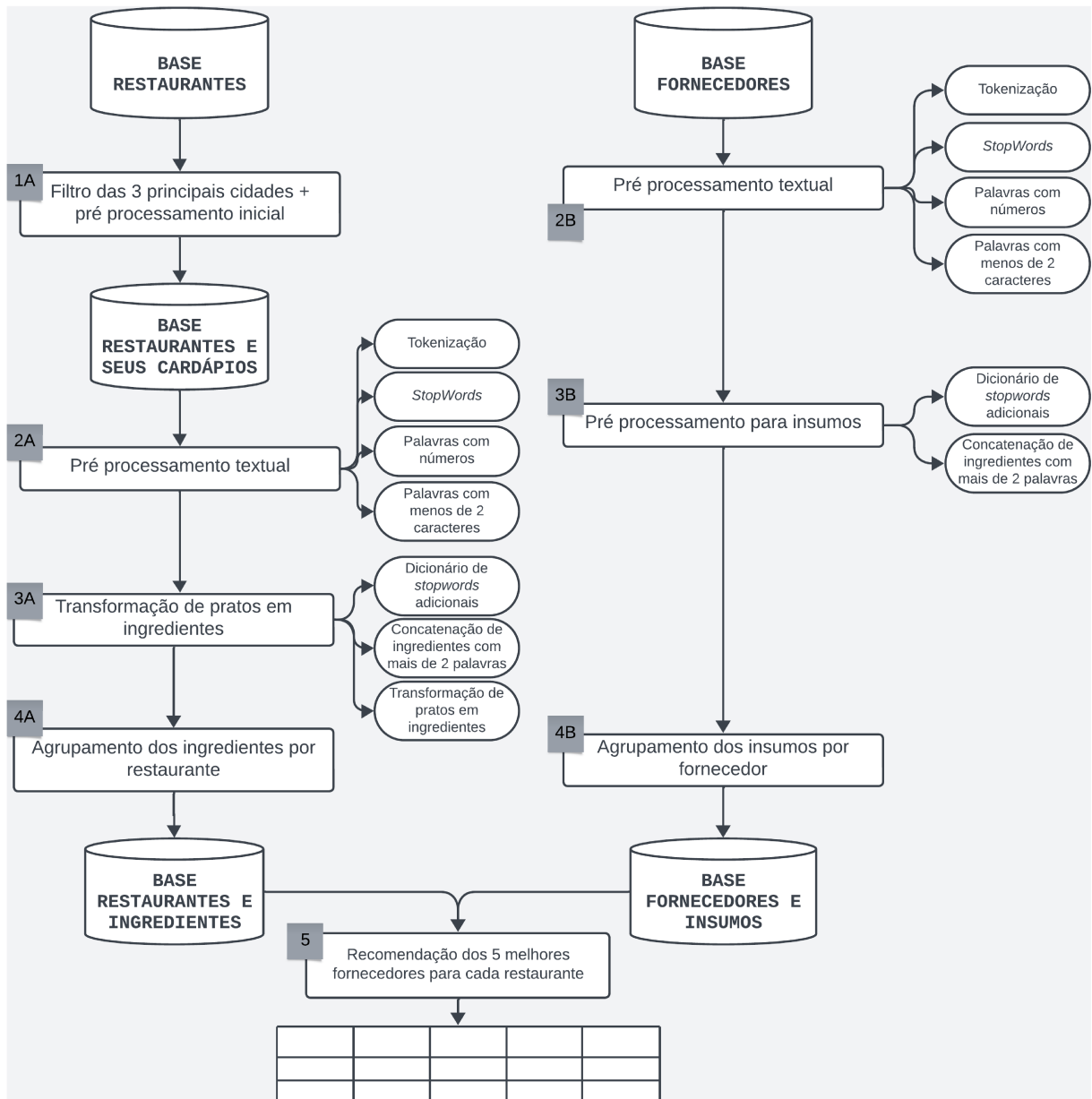


Figura 1 – Etapas do trabalho com a coleta inicial dos dados e seus pré-processamentos aplicados

3.2 Coleta e pré-processamento Inicial

3.2.1 Base dos restaurantes

A primeira etapa realizada neste projeto, como é vista na figura 1, é a coleta dos dados que serão utilizados dos restaurantes. Neste trabalho será utilizada uma base de dados coletada do banco de dados privado da empresa de *delivery* *iFood*.

A extração inicial foi realizada utilizando *SQL* (*Structured Query Language*), uma linguagem para manipulação de dados em bancos de dados. A base extraída contém os restaurantes que já estão cadastrados dentro do *iFood Shop*, plataforma criada pela

empresa *iFood*, que realiza a conexão dos restaurantes com fornecedores de insumos. Foram escolhidas as três principais cidades do negócio: Belo Horizonte, Rio de Janeiro e São Paulo.

Durante a extração foram encontrados erros ortográficos nos nomes das cidades que haviam sido cadastradas na base, sendo assim foram utilizadas técnicas de *SQL* para garantir uma adequação nos nomes dessas cidades, retirando todas as possíveis acentuações e caracteres especiais que possam ter nesses nomes e também os deixando padronizados em letras maiúsculas.

Para fins de exemplo, foram encontradas 24 formas distintas de escrita da cidade de São Paulo. No código foram utilizadas funções via *SQL* para a correção do texto das cidades, como, por exemplo: *translate* e *replace*, para alterarmos os caracteres nos textos, *trim*, para remover espaços em brancos desnecessários e *upper* para deixar os nomes das cidades no padrão em letras maiúsculas.

Após a aplicação de tratamento nos nomes das cidades e realizado o filtro para serem consideradas apenas as três principais (São Paulo, Belo Horizonte e Rio de Janeiro), foram identificados os restaurantes que pertencem à essa base e que serão trabalhados no projeto. Com esses restaurantes identificados foram extraídas as informações de pratos que os mesmos possuem cadastrados dentro do aplicativo de delivery *iFood*. As colunas que foram selecionadas no código possuem descrições, nomes e categorias dos pratos cadastrados.

Os dados que serão trabalhados neste projeto vêm de duas colunas principais: *item_name*, renomeada para "nome_prato" e *item_description*, renomeada para "descricao_prato". Essas duas colunas são inseridas manualmente pelo responsável pelo estabelecimento dentro do aplicativo de *delivery*. O "nome_prato" refere-se ao nome que é cadastrado dentro do aplicativo, e a informação "descricao_prato" é como o estabelecimento descreveu o seu prato no aplicativo, podendo conter informações sobre o prato como o seu peso, ingredientes e características de como foi o modo de preparo, além de, por costume, apresentar diversos adjetivos colocados pelo responsável do estabelecimento. Outras colunas, como a "categoria_aplicativo" e "classificacao_item" também ajudam na classificação dos pratos do restaurante. No entanto, os principais dados para ser criada uma lista dos ingredientes dos estabelecimentos estão dentro das colunas "nome_prato" e "descricao_prato".

As colunas "nome_prato" e "descricao_prato", por serem cadastradas e inseridas manualmente, não possuem um padrão a ser seguido em questões textuais. Assim, tais colunas apresentam diversos caracteres especiais, gírias e erros ortográficos que podem atrapalhar um projeto de mineração de texto. Tais problemas precisam ser tratados de forma eficiente, e a tratativa aqui aplicada inicialmente foi utilizando a linguagem de programação *Spark* e *Python*, retirando todos os possíveis caracteres especiais que possam

atrapalhar a extração de dados. Todas as colunas passam por esse tratamento, no qual são aplicadas funções de pré-processamento, como *unidecode*, para remover caracteres especiais de acentuações e *regex_replace*, para substituir caracteres por outros.

Após todas essas tratativas iniciais feitas, é gerada a base inicial de restaurantes e seus cardápios que este trabalho utiliza. Na tabela 2 são apresentadas todas as colunas extraídas da base original e que serão usadas neste trabalho.

coluna	descricao
merchant_id	identificador único de um restaurante
nome_prato	nome do prato cadastrado no aplicativo
descricao_prato	descricao do prato cadastrado no aplicativo
categoria_aplicativo	categoria no qual o prato se encontra dentro do aplicativo
taxonomia_nome_prato	classificação de bebidas alcólicas ou não alcólicas
classificacao_item	classificação do prato dentre algumas categorias
classificacao_item_parente	classificação que agrupa as classificacoes_itens acima
cidade	cidade onde está presente o restaurante
categoria_restaurante	categoria do tipo de culinária do restaurante

Tabela 2 – Colunas da base inicial de restaurantes e o que cada uma representa

3.2.2 Base dos fornecedores

Durante a coleta de dados da base dos fornecedores não mostrou-se necessária a aplicação de pré-processamentos iniciais como foi realizado no cardápio dos restaurantes.

Por serem trabalhadas as três principais cidades do negócio, conforme descrito na seção 3.2.1, os fornecedores coletados não possuem impedimentos de entrega para os restaurantes presentes na área das cidades escolhidas no projeto. Sendo assim, foram extraídas as informações de todos os fornecedores cadastrados na plataforma *iFood Shop* e seus respectivos insumos que são oferecidos na plataforma.

A coleta de dados dos fornecedores resulta em uma tabela com as colunas apresentadas e explicadas presente na tabela 3.

coluna	descricao
seller_id	identificador único de um fornecedor
sku	identificador único de um insumo
nome_insumo	descricao do produto cadastrado na plataforma
categoria	categoria no qual o produto corresponde

Tabela 3 – Colunas da base inicial de fornecedores e o que cada uma representa

3.3 Pré-processamento da base textual

3.3.1 Base dos restaurantes

O pré-processamento textual dos restaurantes ao qual se refere a etapa 2A da imagem 1 se divide em etapas e regras para que sejam excluídas palavras iniciais consideradas desnecessárias para a base de cardápio dos restaurantes.

A etapa de *tokenização* é necessária por se tratar de textos com diversas palavras, dividindo então os nomes e descrições dos pratos em palavras individuais que serão avaliadas e contadas individualmente para as próximas regras aplicadas.

Com a tokenização realizada são retiradas da base todas as palavras que estejam presentes na lista de *stopwords* padrão da biblioteca da linguagem *Python NLTK*. Dentro dessa lista, no entanto, foi retirada a palavra "de" por representar uma importante tarefa neste trabalho.

A palavra "de" é muito utilizada para concatenar ingredientes e produtos que são escritos com mais de uma palavra. Ao apresentar o texto, por exemplo, "creme de leite" o código precisa entender que as três palavras se referem a um ingrediente único. Caso seja retirada a palavra "de" esse ingrediente seria representado como "creme leite" apresentando um significado diferente do original.

A segunda regra aplicada nesta etapa de pré-processamento textual é a identificação e remoção de palavras com números. Existem diversos nomes e descrições de pratos que possuem números representando quantidades, peso e tamanho dos mesmos, e tais informações são desnecessárias para o objetivo deste trabalho. Sendo assim, textos que apresentem, por exemplo, "500ml", são retirados da base final.

A última regra de remoção aplicada nesta etapa se dá pelas palavras que possuam 2 ou menos caracteres. Por serem colunas inseridas manualmente foram encontrados diversos erros ortográficos em que apenas uma ou duas letras estavam presentes no nome e descrição dos pratos não apresentando significado algum para a base. Também foram identificadas diversas palavras como "ml" e "gr" representando o peso de produtos, que também foram retiradas.

3.3.2 Base dos fornecedores

Na etapa 2B, etapa de pré processamento inicial na base dos fornecedores, foi realizado um processo similar ao que já foi apresentado na seção 3.3.1. Para a base de fornecedores também foi realizada a *tokenização* dos textos e posteriormente foram retiradas as palavras que se encaixem dentro da lista de *stopwords* padrões da biblioteca do *Python*.

Essa base de fornecedores possuía uma quantidade importante de palavras com caracteres numéricos no meio delas. Isso se dá devido à necessidade do fornecedor informar ao usuário (o restaurante) a quantidade exata de produtos que o estabelecimento está comprando assim como o peso correto dos insumos fornecidos. Sendo assim, foram retiradas todas as palavras que possuem caracteres numéricos.

A retirada de palavras com dois ou menos caracteres também ocorreu e, da mesma forma que a base de restaurantes, a palavra "de" foi retirada dessas regras visto que será utilizada como uma palavra importante para identificar insumos posteriormente.

3.4 Transformação de pratos em ingredientes

Após a aplicação do conjunto de técnicas de pré-processamento detalhados na subseção 3.3.1, são realizadas três novas fases de pré-processamento com o objetivo final de transformação de pratos comuns na base dos cardápios em ingredientes específicos que o compõe, como é visto na figura 1.

Inicialmente é realizada uma contagem de frequência em toda a base das palavras presentes nos textos das colunas "descricao_prato" e "nome_prato". Essa contagem de frequência é utilizada para criar-se um dicionário adicional de *stopwords* e ter uma ideia de principais pratos cadastrados na base.

Esse novo dicionário de *stopwords* é composto por palavras que possuam uma baixa frequência em toda a base e palavras que foram manualmente coletadas que possuam uma frequência relevante mas que não sejam importantes para o projeto.

Foram encontradas 199.107 palavras distintas em toda a base, e apenas 3.426 possuem uma frequência maior que 1.500 vezes. Todas as palavras que possuam uma frequência menor que 1.500 foram inseridas no novo dicionário de *stopwords* por não apresentarem sentido para a identificação de ingredientes específicos.

Além das palavras com baixa frequência, foram identificadas diversas palavras que possuem uma alta frequência e que poderiam ser retiradas dos textos para que não afetem a recomendação criada entre restaurantes e fornecedores. Palavras como "combo", "acompanha" e "lata" mostraram uma frequência alta na base, além de palavras que demonstram adjetivos para os pratos como "delicioso" e "tradicional". Essas palavras foram adicionadas ao dicionário de *stopwords* adicionais e todas serão removidas dos textos presentes nas colunas "nome_prato" e "descricao_prato".

A segunda fase da etapa 3A é a concatenação de ingredientes com 2 ou mais palavras. Utilizando a listagem de frequência de palavras que foi criada no início desta etapa foram identificadas palavras que, sozinhas, podem não apresentar o seu real significado. A segunda palavra com maior frequência na base, desconsiderando a palavra "de", é

"molho", e foram identificadas também palavras com frequência alta como, por exemplo, "pao", "creme", "leite" e "condensado".

Para resolver este problema foram realizadas concatenações de palavras a partir de padrões que são encontrados na base. Alguns desses padrões podem ser vistos na Tabela 4.

texto_original	texto_modificado
'coca cola'	'cocacola'
'molho de tomate'	'molhodetomate'
'creme de leite'	'cremedeleite'
'batata frita'	'batatafrita'
'batata doce'	'batatadoce'
'pao australiano'	'paoaustraliano'
'farinha de rosca'	'farinhaderosca'

Tabela 4 – Tabela demonstrando concatenação de palavras

Essa fase de concatenação de palavras permite serem identificados ingredientes que poderiam apresentar sentidos diferentes do que o texto original apresenta, e essa má interpretação do texto poderia apresentar um erro de recomendação quando for realizada a junção com a base de fornecedores.

Como um exemplo para a motivação desta concatenação temos que, caso a palavra "batata", que no contexto de um cardápio específico de um restaurante especializado em sanduíches represente "batata frita", fosse deixada sozinha na base, ela poderia ser relacionada com a palavra "batata" em um fornecedor que, na realidade, represente "batata doce", e não o produto "batata_frita" ao qual o restaurante necessita, levando assim a uma má interpretação do cardápio e resultando em uma recomendação errônea baseada no duplo sentido de uma palavra única.

A última aplicação de pré-processamento na etapa 3A é a transformação de pratos em ingredientes. Utilizando as palavras mais frequentes na base foram identificados pratos que possuem uma alta taxa de aparição nos diversos cardápios dos restaurantes cadastrados na base. Pratos como: *stroganoff*, sanduíche, pizza, escondidinho, *hot roll*, entre outros, aparecem com frequência na base por serem os pratos mais famosos dentro do aplicativo de *delivery*.

Estes pratos fornecem importantes informações sobre quais os insumo que um restaurante pode necessitar para o preparo de seu cardápio. Desta forma, todos os textos que apresentem algum destes principais pratos na base são transformados e quebrados em palavras que denotem os ingredientes específicos relacionados ao que o restaurante está oferecendo em seu cardápio.

Essa transformação de textos representando pratos em seus ingredientes específicos

proporciona um aumento significativo na quantidade de vezes que alguns ingredientes aparecem na base final.

Este aumento na frequência de ingredientes importantes resulta em um sistema de recomendação mais robusto, podendo realizar conexões mais assertivas entre fornecedores e restaurantes. Um fornecedor, por exemplo, não irá apresentar em seu catálogo de insumos um produto específico chamado *temaki*, porém ele pode apresentar os ingredientes específicos deste produto, como: alga, arroz, *cream cheese*, salmão, pepino, atum, molho *teriyaki* e molho *shoyu*.

Para a base de fornecedores, detalhando a etapa 3B da Figura 1, são realizadas as duas primeiras fases apresentadas nesta subseção. Desta forma, os insumos oferecidos pelos fornecedores são concatenados, para representarem ingredientes únicos, como demonstrado na Tabela 4, e as *stopwords* adicionais também são retiradas de seus textos. Para os fornecedores não é necessária a etapa de transformação de pratos em ingredientes dado que eles já possuem em seus textos o insumo final que será utilizado pelos restaurantes, e não o prato específico.

3.5 Agrupamento de ingredientes e insumos

Após a aplicação do conjunto de técnicas de pré-processamento nos cardápios dos restaurantes e nos insumos dos fornecedores descritos na Seção 3.4 é realizado o agrupamento de todos os ingredientes e insumos por cada restaurante ou fornecedor único.

O resultado das etapas 3A e 3B, da Figura 1, proporciona que cada prato do restaurante tenha sido descrito como seus próprios ingredientes. Dessa forma é necessário o agrupamento de todas as palavras, de todos os pratos, que cada restaurante possui cadastrado em seu cardápio.

Dessa forma, cada *merchant_id*, apresentado na tabela 2, vai possuir uma coluna específica chamada de "ingredientes", que apresenta todas as palavras distintas que foram encontradas nas colunas "descricao_prato" e "nome_prato" após a aplicação dos pré-processamentos, resultando assim em todos os ingredientes que aquele restaurante utiliza em todo o seu cardápio.

A mesma estratégia é utilizada na base de fornecedores. Como visto na 3, cada linha presente na base apresentava um insumo específico oferecido pelo fornecedor no sistema. Desta forma, é realizado um agrupamento a nível de "seller_id" com todas as palavras que representem os insumos que aquele fornecedor específico possui em sistema para venda.

Ao final desses agrupamentos realizados são obtidas duas tabelas, no modelo das Tabelas 5 e 6, representando a "Base Restaurantes e Ingredientes" e "Base Fornecedores

e Insumos" da Figura 1.

seller_id	insumos
seller_1	insumos_seller_1
seller_2	insumos_seller_2

Tabela 5 – Tabela final dos fornecedores

merchant_id	ingredientes
restaurante_1	ingredientes_restaurante_1
restaurante_2	ingredientes_restaurante_2

Tabela 6 – Tabela final dos restaurantes

3.6 Sistema de recomendação

Com as duas bases criadas ao final da etapa de agrupamento vistas na seção 3.5 é criado um método para avaliar quantos ingredientes e insumos existem em comum entre um restaurante e todos os fornecedores cadastrados na base.

É realizada uma contagem de quantos ingredientes um restaurante possui em seu cardápio e quantos desses ingredientes estão presentes em cada base de insumos para cada fornecedor cadastrado na base.

Sendo assim, todos os restaurantes vão ter os seus respectivos ingredientes verificados na base de insumos de cada fornecedor. Com essa relação é realizado um cálculo que consiste na divisão de quantos ingredientes foram encontrados para aquele fornecedor pelo tanto de ingredientes total que o restaurante possui em seu cardápio.

Esse cálculo de percentual é a métrica que será utilizada neste projeto para ser determinado o quanto um fornecedor é compatível ou não com algum restaurante, sendo que ela representa o quanto aquele fornecedor específico consegue atingir de oferta para o cardápio de um estabelecimento específico.

Ao final da avaliação de cada fornecedor para um restaurante específico são selecionados os 5 melhores fornecedores para aquele restaurante, sendo eles determinados pelos maiores percentuais de insumos e ingredientes em comum, resultando na base de recomendação vista na figura 1 após a etapa 5.

Com os 5 melhores fornecedores para cada restaurante é possível realizar comunicações mais assertivas no que se refere às técnicas de *CRM*, finalizando assim o objetivo deste trabalho.

3.7 Considerações Finais

Neste capítulo foram descritas todas as formas de pré-processamento textuais necessárias, assim como foram criados dicionários de apoio para a limpeza dos dados e um dicionário extra que transforma pratos em ingredientes para, por fim, ser realizada uma

junção entre duas bases distintas com o objetivo de criação do sistema de recomendação inicial.

Como descrito na seção 2.8, é visto que esses sistemas de recomendação se baseiam em mostrar para o usuário a melhor opção possível a ele baseada em suas próprias preferências, com o objetivo de manter este usuário dentro de alguma plataforma. Este projeto utiliza os ingredientes que o restaurante possui para criar um sistema de recomendação de fornecedor e demonstrar com clareza de que o restaurante pode utilizar da plataforma *iFood Shop* para comprar boa parte de seus insumos, mantendo assim o restaurante dentro da plataforma *iFood* como um todo.

No próximo capítulo serão apresentadas as volumetrias utilizadas neste trabalho, assim como os resultados que foram obtidos, desde os pré-processamentos até o cálculo da métrica que relaciona insumos e ingredientes.

4 Resultados

Neste capítulo são mostrados os resultados obtidos nas etapas de coleta, pré-processamento e recomendação realizados neste projeto, evidenciando os resultados obtidos por cada etapa da Figura 1.

Na seção 4.1 é mostrado o tamanho da base usada neste projeto e os pré-processamentos iniciais realizados para a extração da primeira base de restaurantes e fornecedores.

Em 4.2 são mostrados os resultados dos pré-processamentos específicos aplicados na base de dados, os quais consideraram a frequência de ocorrência de palavras na base.

A seção 4.3 mostra os resultados dos textos dos cardápios após os pré-processamentos realizados com o intuito de transformar pratos dos restaurantes em ingredientes específicos.

Por fim, os resultados do sistema de recomendação criado são detalhados na seção 4.4. A seção 4.5 mostra um breve resumo sobre todos os assuntos trabalhados neste capítulo.

4.1 Coleta de dados e volumetria da base

Esta seção apresenta o resultado da coleta dos dados e a volumetria obtida da base de restaurantes e fornecedores.

4.1.1 Base dos restaurantes

Como descrito na seção 3.2, foram coletados todos os restaurantes presentes nas três principais cidades para o negócio: São Paulo, Belo Horizonte e Rio de Janeiro e que estavam cadastrados na plataforma do *iFood Shop*.

Alguns exemplos dos resultados das tratativas feitas nos nomes das cidades podem ser vistos na Tabela 7.

Na Figura 2 é visto um prato cadastrado dentro do aplicativo utilizado aqui como exemplo. Na imagem são vistas as duas colunas que foram utilizadas neste trabalho: `nome_prato`, sendo "MEXIDÃO DO REI (2 PESSOAS)", e `descricao_prato`, sendo todo o texto abaixo do nome do prato.

Após feito o pré processamento inicial nesse texto, presente na etapa 1A da Figura 1, são obtidas as duas colunas na base sem acentuações e caracteres especiais, além de

texto_original	texto_modificado
'SAo pauLO'	'SAO PAULO'
'saõ paulo'	'SAO PAULO'
'blo horizonte'	'BELO HORIZONTE'
'BELO horizonte'	'BELO HORIZONTE'
'rio d janeiro'	'RIO DE JANEIRO'
'rj'	'RIO DE JANEIRO'

Tabela 7 – Tabela demonstrando tratativa nos nomes das cidades

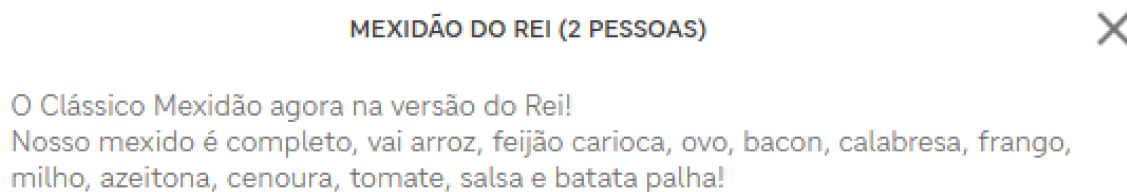


Figura 2 – Um prato cadastrado no aplicativo

todas as letras minúsculas. O resultado desse pré processamento, quando aplicado na Figura 2 é apresentado na Tabela 8.

Nome do Prato	Descrição do Prato
mexidao do rei 2 pessoas	o classico mexidao agora na versao do rei nosso mexido e completo vai arroz feijao carioca ovo bacon calabresa frango milho azeitona cenoura tomate salsa e batata palha

Tabela 8 – Prato de exemplo após o pré processamento inicial

A base extraída para utilização neste projeto possui a volumetria apresentada na Tabela 9. Por cidades foram obtidos: 98.579 restaurantes em São Paulo, 55.995 restaurantes no Rio de Janeiro e 22.227 restaurantes em Belo Horizonte. Em média, cada restaurante que foi selecionado para compor a base possui, cadastrado no aplicativo de delivery *iFood*, 34 pratos distintos em seus cardápios (contagem feita pela coluna "nome_prato") e 28 descrições de pratos distintas.

Total de linhas	6.293.406
Restaurantes distintos	176.781
Nomes de pratos distintos	1.873.152
Descrições de pratos distintos	2.424.482
Categorias de pratos cadastradas no aplicativo	188.974
Categorias de restaurantes	60
Cidades	3

Tabela 9 – Volumetria da base de restaurantes

A tabela 10 apresenta os principais tipos de culinária presentes na base de dados do projeto. Esses principais tipos de culinária foram escolhidos a partir dos dados privados de faturamento que o *iFood Shop* atualmente apresenta, sendo essas oito culinárias as mais representativas para o negócio. O total de restaurantes cadastrados em cada categoria e o nome de prato mais frequente dentro daquela categoria. O nome de prato presente na Tabela é o nome tratado apenas com os pré-processamentos iniciais realizados na etapa 1A da Figura 1 para a retirada de acentuações e caracteres especiais.

Tipo de culinária	Total de restaurantes	Nome de prato mais frequente
LANCHES	38.140	x bacon
COMIDA BRASILEIRA	32.697	file de frango
DOCES & BOLOS	21.499	brigadeiro
PIZZA	13.960	grande 8 pedacos
BEBIDAS	9.601	fanta laranja
ACAI	8.248	acai 500ml
MARMITA	7.815	file de frango
COMIDA JAPONESA	7.045	temaki california

Tabela 10 – Volumetria e principais nomes de pratos presentes por tipo de categoria do restaurante

4.1.2 Base dos fornecedores

Na Figura 3 é apresentado um exemplo de um dos produtos cadastrados no *iFood Shop*.

Açúcar Refinado Especial União Caixa com 1.0 Kg

R\$ 4,35/kg

1 kilo
Valor total: R\$ 4,35

[Comparar preços](#)

Adicionar ao carrinho

[Salvar na lista](#)

Informações do fornecedor

Vendido por **Extra Atacado**

Pedido Mínimo R\$ 400,00

Informações de entrega

Figura 3 – Um insumo do fornecedor cadastrado no aplicativo

Este produto, oferecido pelo fornecedor "Extra Atacado", passa por tratativas de pré-processamento no nome do produto, resultando em um produto com o nome: "acucar refinado especial uniao caixa com 1 0 kg".

A volumetria da base de fornecedores é apresentada na Tabela 11.

Total de linhas	1.106.871
Fornecedores distintos	2.014
Total de insumos distintos	218.766
Média de insumos por fornecedor	270

Tabela 11 – Volumetria da base de fornecedores

4.2 Resultados pré-processamento textual

Alguns exemplos da remoção de *stopwords*, palavras numéricas e palavras com menos de 2 caracteres, e como ela afeta as descrições dos pratos cadastrados, podem ser vistos na Tabela 12. Obteve-se uma base expressivamente mais limpa após a aplicação de remoção das palavras que se encaixem nas regras determinadas.

A base que inicialmente possuía um total de 241.928 palavras distintas, reunindo todas as palavras presentes nas colunas "nome_prato" e "descricao_prato", passou a apresentar um total de 199.108 palavras, uma queda significativa de **17%** de palavras que não iriam apresentar funções para a criação do sistema de recomendação.

Descrição do Prato antiga	Descrição do Prato sem <i>stopwords</i>
com uma casquinha de chocolate por fora e super macio por dentro gotas de chocolate ao leite derramadas pela massa esse brownie e de comer de joelhos	casquinha de chocolate super macio dentro gotas de chocolate leite derramadas massa brownie de comer de joelhos
combo familia 2 bife a milanesa 2 bife a parmegiana coca 2l gratis	combo familia bife milanesa bife parmegiana coca gratis
pao smash de carne 90gr queijo cebola caramelizada e maionese house ponto da carne unico bem passado	pao smash de carne queijo cebola caramelizada maionese house ponto carne unico bem passado
deliciosa e succulenta feijoada caprichada conhecida ha anos como a melhor da regioao acompanha couve farofa e torresmo	deliciosa succulenta feijoada caprichada conhecida anos melhor regioao acompanha couve farofa torresmo

Tabela 12 – Exemplos de descrições de pratos após a remoção de *stopwords*

Os mesmos tratamentos foram aplicados na base de fornecedores. Esta base apresentava maior quantidade de palavras com caracteres numéricos, dado que os fornecedores precisam explicitar aos restaurantes a quantidade de itens que estão sendo vendidos, assim como o peso exato de seus produtos. Algumas dessas palavras, e a quantidade de vezes que elas aparecem na base, podem ser vistas na Tabela 13.

A base inicial dos fornecedores possuía 43.644 palavras distintas na coluna de nome dos produtos e, após a aplicação de remoção dessas palavras iniciais, a base ficou com 25.897 palavras distintas, uma redução de aproximadamente **41%**, produzindo assim uma

palavra	frequência
10	376.438
gr	174.204
kg	141.548
500	56.196
cm	35718
120	35.256

Tabela 13 – Tabela com palavras frequentes que foram removidas da base de fornecedores base de insumos de fornecedores mais limpa para a realização de recomendação para os restaurantes.

4.3 Resultados da transformação de pratos em ingredientes

Após a remoção de algumas das principais palavras desnecessárias da base de restaurantes, conforme regras explicadas na seção 4.2, foi realizada uma nova contagem de frequência das palavras a partir dos textos presentes nas colunas "nome_prato" e "descricao_prato". Algumas das palavras mais frequentes podem ser vistas na Tabela 14.

palavra	frequencia
queijo	813.939
molho	754.959
frango	747.528
batata	711.214
chocolate	665.269
acompanha	354.456
sabor	300.445
combo	298.249
especial	280.411

Tabela 14 – Contagem de frequência de palavras na base

Com essa contagem, foram identificados novos pré-processamentos que se mostraram necessários na base. Algumas palavras, como "acompanha", "sabor", "combo" e "especial" aparecem em uma frequência extremamente alta, e tais palavras não apresentam importância para o sistema de recomendação que será criado, porém elas não fazem parte das regras que foram aplicadas na seção 4.2.

Entendendo a necessidade de remover mais palavras, foi criado manualmente um dicionário adicional de *stopwords*, com palavras identificadas com uma alta frequência de ocorrência na base, e também foram adicionadas a esse dicionário palavras com frequência extremamente baixa.

Essa remoção de *stopwords* adicionais proporciona uma base com tamanho menor,

visto que palavras como "acompanha" apareciam um total de 354.456 vezes, e também impede que tais palavras sejam reconhecidas posteriormente dentro do sistema de recomendação, quando for realizada a recomendação para os restaurantes a partir dos pratos presentes no seu cardápio.

Além das palavras manualmente coletadas com uma frequência alta, foram também desconsideradas todas as palavras que possuem uma frequência menor que 1.500 vezes na base por não demonstrarem ingredientes ou pratos específicos para a base. De um total de 199.107 palavras distintas presentes nos textos das colunas "descricao_prato" e "nome_prato", apenas 3.427 apresentam uma frequência maior que 1.500.

Por serem dados inseridos manualmente no aplicativo de *delivery*, muitas palavras presentes nos textos dos nomes e descrições dos pratos não possuem relevância para o projeto. Cerca de 150.000 palavras presentes na base possuem uma aparição de 10 vezes para menos, e essa frequência é extremamente baixa para uma base deste tamanho, como foi verificado na Tabela de volumetria da seção 4.1.1.

O resultado aplicado dessa remoção adicional de *stopwords* podem ser conferidos com alguns exemplos na Tabela 15. Os pratos colocados no exemplo são os mesmos utilizados na Tabela 12. Ao total, 196.437 palavras entraram no dicionário adicional de *stopwords*, e elas foram disponibilizadas via *GitHub*¹.

Descrição do Prato Sem <i>stopwords</i>	Descrição do Prato sem <i>stopwords</i> adicionais
casquinha de chocolate super macio por dentro gotas de chocolate ao leite derramadas pela massa esse brownie e de comer de joelho	casquinha chocolate gotas chocolate leite brownie
combo familia 2 bife a milanesa 2 bife a parmegiana coca 2l gratis	bife milanesa bife parmegiana coca
pao smash de carne queijo cebola caramelizada maionese house ponto carne unico bem passado	pao smash carne queijo cebola caramelizada maionese house carne passado

Tabela 15 – Exemplos de descrições de pratos após a remoção de *stopwords* adicionais

A próxima fase da etapa 3A, vista na Figura 1, é a concatenação de ingredientes com mais de duas palavras. Algumas concatenações foram feitas para ter em uma única palavra, produtos, ou ingredientes, que são representados por mais de uma palavra.

Todo texto que apresentasse, por exemplo, a sequência "pao de hamburger", foi concatenado para apresentar, em apenas uma palavra, "paodehamburger". Também foram feitas adequações de produtos que possam ser representados de duas maneiras distintas, como é o exemplo de "cachorro quente", que usualmente pode ser escrito também como "hot dog".

¹ https://github.com/Hesloan/TCC_Hesloan/blob/main/dicionario_stopwords.txt

Alguns dos exemplos de concatenação realizados nesta base podem ser vistos na Tabela 16.

sequencia de palavras	palavra unica
pao de hamburguer	paodehamburguer
creme de leite	cremedeleite
leite condensado	leitecondensado
batata frita	batatafrita
coca cola	cocacola
farinha de trigo	farinhadetrigo
pao brioche	paobrioche
molho de tomate	molhodetomate
carne moída	carnemoida

Tabela 16 – Exemplos de concatenação de palavras feitas

Após realizada essa concatenação, observou-se uma mudança importante na frequência das palavras que anteriormente haviam sido analisadas. A palavra concatenada "batatafrita", por exemplo, se mostrou extremamente relevante para a base, aparecendo um total de 486.313 vezes. Outros exemplos, como "molhodetomate", "molhoteriyaki", "creamcheese", "paobrioche" e "cocacola" demonstraram uma grande participação de frequência nos cardápios dos restaurantes.

Palavras frequentes que representem métodos de preparo como "salteado", "grelhado", "refogado" e "selado" passaram por mudanças para apresentarem ingredientes específicos como "óleo azeite manteiga". Com essa mudança de palavras, observou-se um aumento significativo na frequência dessas palavras que representam os ingredientes desses métodos de preparo, como é o caso da palavra "grelhado" e os resultados obtidos podem ser vistos na Tabela 17.

palavra	contagem_antes	contagem_depois
oleo	38.521	420.094
manteiga	77.539	459.112
azeite	57.633	439.206

Tabela 17 – Contagem de vezes em que as palavras óleo, manteiga e azeite aparecem antes e depois do pré-processamento realizado

Foram realizadas também transformações em palavras que apresentaram erros ortográficos mas que demonstraram frequências consideráveis nas contagens feitas, assim como palavras que necessitavam de adequações, como foi observado no caso de "catchup", que aparecia 15.609 vezes na base, enquanto "ketchup" aparecia 71.076 vezes, e este mesmo caso se aplicou para as diversas formas de escrita da palavra "mussarela", sendo necessário um adequamento de texto toda vez que as palavras "muzzarela", "muzarela" e "mucarela" apareciam.

Por fim, alguns pratos frequentes na base foram transformados para apresentarem seus ingredientes específicos. Alguns exemplos dessas transformações podem ser visualizados na Tabela 18. As palavras em destaque na tabela representam onde foram feitas as transformações.

descricao_prato_original	descricao_prato_transformada
strogonoff de frango cremoso, acompanha batata palha e uma coca 350ml	peitodefrango ketchup cebola alho mostarda molhodetomate cremedeleite cogumelo batatapalha cocacola
prato contém 10 unidades hot roll, acompanha 2 molhos teriyaki	oleo alga arroz creamcheese molhoteriyaki salmao molhoteriyaki
sanduche com 2 smash, acompanha batata frita e 2 saches de catchup	carnemoida patinho acem paodehamburger paobrioche hamburger queijo cheddar batatafrita ketchup
escondidinho de frango, acompanha arroz e batata palha	peitodefrango mussarela azeite manteiga requeijao arroz batatapalha
hot dog duplo com 2 salsichas acompanha 1 coca cola	paodehotdog salsicha milho presunto mussarela molhodetomate ketchup mostarda maionese cocacola

Tabela 18 – Transformação de pratos em ingredientes específicos

A aplicação de todas essas fases de pré-processamento proporcionaram um aumento significativo nas palavras que possuam significados de ingredientes para a base. O termo concatenado previamente "peitodefrango", um ingrediente que faz parte de boa parte dos restaurantes, aparecia na base inteira um total de 149.905 vezes. Após a transformação de pratos em ingredientes, o termo "peitodefrango" apareceu 426.982 vezes, e esse aumento proporciona que cada vez mais restaurantes possuam em seu cardápio final de ingredientes o item "peitodefrango", elevando o potencial de uma recomendação correta dos fornecedores.

Na Tabela 19 é demonstrado como ficaram todas as palavras dos cardápios de alguns restaurantes antes e após a aplicação de todas as etapas de pré-processamento.

4.4 Recomendação de fornecedores

Os resultados das recomendações de fornecedores encontrados se mostraram positivos. Em média, considerando o *top 1* fornecedor de cada restaurante, **87%** dos ingredientes do cardápio dos restaurantes foram ofertados por esses "*top1_fornecedores*".

A cidade que obteve uma menor média de ingredientes identificados dentro dos insumos dos fornecedores foi o Rio de Janeiro, com uma média de 84% de ingredientes

palavras_cardapio_inicial	palavras_cardapio_final
no cenoura vulcaozinho meio bolo com ninho chocolate piscina laterais mousse uma de limao e maravilhosa recheio recheado pool chocolatudo cake massa cobertura	chocolate cenoura chocolatudo cake limao leiteninho mousse bolo
os crispy dupla verde caramelizada pao bacon bovina man tomate batata alface jetsons cheddar flinstones lata looney rings fanta ovo carne 350 de antartica he tunes ml laranja hamburguer roxa cebola popeye barbecue queijo 350ml cada australiano media artesanal brioche duplo coca guarana bovino onion molho 150g	queijo carne acucarmascavo cebola bovina tomate guarana ovo shoyu cocacola barbecue acucar pao bovino alface farinhadetrigo batata antartica bacon paodebrioche azeite paoaustraliano cheddar manteiga fanta hamburguer oleo laranja
seco gas original 750ml antarctica broto boi grande cerveja garrafa 5l pequena 18 chalise lata sangue pedacos cola maiores cerv branco dolly itaipava refrigeranteschweppes de suave sabores refrigerante 8 skol tinto lt vinho 2 agua 1 crystal 350ml citrus 2l produto anos 4 para coca 3 sem guarana 500ml	cerveja crystal dolly itaipava antarctica broto refrigerante boi skol tinto guarana vinho agua cocacola

Tabela 19 – Exemplos de cardápios de três restaurantes distintos antes e depois dos pré processamentos.

encontrados. São Paulo obteve uma média de 87,1% e Belo Horizonte foi a cidade com a maior média de ingredientes encontrados por restaurante, com 87,5%.

Na Tabela 20 são encontradas as médias por cada tipo de culinária considerando as principais categorias cadastradas na base.

A categoria "comida japonesa" apresentou uma média menor do que as outras categorias. Também foram encontrados bem menos fornecedores diferentes para este tipo de restaurante. Isso pode ser explicado por conta da complexidade de alguns pratos estrangeiros de serem identificados e transformados corretamente em seus ingredientes específicos assim como uma menor quantidade de fornecedores específicos para este tipo de culinária. Para este tipo de culinária foram transformados os principais pratos de seus restaurantes, como *hot roll*, *temaki*, *sunomono* e *sushi's* no geral.

Já a categoria "bebidas" se encaixa em um nível de dificuldade oposto à categoria japonesa. Por serem restaurantes cadastrados com o objetivo de venderem bebidas, não

Tipo de culinária	Média de recomendação	Quantidade de fornecedores distintos encontrados
LANCHES	88,6%	134
COMIDA BRASILEIRA	89%	145
DOCES & BOLOS	80%	134
PIZZA	90%	94
BEBIDAS	92%	103
ACAI	81%	109
MARMITA	89%	89
COMIDA JAPONESA	77%	49

Tabela 20 – Volumetria e principais nomes de pratos presentes por tipo de categoria do restaurante

existem, nesses estabelecimentos, pratos dos quais sejam necessários um tipo de tratamento específico. As concatenações realizadas na etapa 3A da Figura 1 se mostraram suficientes para este tipo de restaurante.

Este tipo de resultado pode proporcionar uma comunicação muito mais assertiva vindo da equipe de *CRM* do *iFood Shop* visto que cada restaurante possui agora o seu fornecedor com maior potencial de compra dada a relação entre produtos do restaurante e insumos do fornecedor.

4.5 Considerações finais

Os resultados deste projeto se mostraram significativos para o que foi proposto inicialmente, relacionando os restaurantes e seus cardápios cadastrados no aplicativo do *iFood* com os melhores fornecedores possíveis dentro da plataforma *iFood Shop*.

As diversas fases de pré-processamento aplicadas neste projeto se mostraram relevantes para todo o processo de transformar os cardápios em ingredientes. Era esperado inicialmente também que alguns tipos de restaurantes específicos, como os restaurantes de comida japonesa, demonstrassem uma dificuldade a mais por conta de seus pratos e ingredientes específicos. Essa preocupação se mostrou pertinente ao analisar os resultados, porém, mesmo com essa dificuldade extra, o percentual de ingredientes desses restaurantes que foram encontrados nos fornecedores se mostrou de uma alta utilidade para serem aproveitados pela equipe do *iFood Shop*.

5 Conclusão

Este trabalho possui como objetivo principal a criação de um sistema de recomendação de fornecedores para restaurantes baseado em uma relação entre os insumos oferecidos pelos fornecedores e os ingredientes que os restaurantes utilizam dados os cardápios de pratos que os mesmos apresentam no aplicativo de *delivery iFood*, e os resultados obtidos após as diversas etapas de trabalho se mostraram importantes.

A partir dos dados coletados dos cardápios cadastrados no *iFood* foram realizadas diversas etapas de pré-processamentos para que os textos apresentassem apenas o que o sistema de recomendação necessitava, ou seja, foram diversas etapas para que os pratos cadastrados no aplicativo se transformassem em ingredientes específicos que possam ser usados naqueles pratos.

A criação do sistema de recomendação tem como objetivo enriquecer as comunicações que a equipe de *CRM* da plataforma *iFood Shop* possa fazer com seus clientes. O resultado obtido pelo sistema de recomendação mostrou-se útil para que a equipe de *CRM* pudesse utilizar em suas comunicações. O percentual de ingredientes que os restaurantes possuem que foram encontrados na listagem dos insumos dos fornecedores foi de, em média, 87%, apresentando um número até maior para tipos de restaurantes específicos, como *pizzarias* e restaurantes focados em entrega de bebidas.

Toda a idéia por trás do sistema de recomendação proposto é para que pudessem ser mostrados de forma mais clara os melhores fornecedores dentro da plataforma do *iFood Shop* para cada restaurante específico. A partir do cálculo de quantos ingredientes um fornecedor atende ao cardápio do restaurante, pode-se dizer que esse fornecedor é de importância significativa para o estabelecimento, para que ele possa comprar o máximo de ingredientes que ele necessita apenas de um fornecedor específico.

5.1 Contribuições

As contribuições deste trabalho foram:

1. Criação de um dicionário de *stopwords* referente à cardápios de restaurantes
2. Criação de regras para a concatenação e transformação de pratos em ingredientes específicos
3. Construção de um sistema de recomendação baseado no quanto um fornecedor pode contribuir para os produtos que um restaurante necessita

4. Indicação mais assertiva de fornecedores para as comunicações de *CRM* da equipe do *iFood Shop*

5.2 Trabalhos futuros

Foram utilizadas diversas regras de pré-processamento para a construção deste sistema de recomendação, porém algumas delas podem ser aprimoradas para que não seja necessário um trabalho manual em cima destas etapas. A transformação de pratos em ingredientes, por exemplo, poderia ser feita de uma forma mais automatizada pensando em coletar dados desses pratos de *sites* especializados em receitas. Dessa forma, quando o texto apresenta algum tipo de nome de prato, o código procuraria diretamente nos *sites* quais os ingredientes que aquele prato específico apresenta.

Para o sistema de recomendação proposto aqui foi utilizada uma regra simples analisando se a palavra que representa um ingrediente do restaurante foi ou não encontrada na listagem de insumos dos fornecedores. Uma forma de aprimoramento desta técnica é a utilização de algoritmos que calculem a similaridade entre duas palavras, e determinar o quão duas palavras podem ser similares para que a recomendação possua um % de acerto maior.

Um terceiro ponto a ser avaliado para aprimoramento é o cálculo mais complexo de recomendação de fornecedores utilizando não só o número de ingredientes e insumos em comum, mas também avaliando o valor de pedido mínimo do fornecedor, o valor do frete do fornecedor para aquele restaurante específico e o tempo de entrega. Essas três métricas podem afetar o quanto um estabelecimento vai ou não querer comprar de um fornecedor. Um fornecedor que apresente uma relação de 60% do cardápio do restaurante em sua listagem de insumos pode ser mais atrativo para o restaurante do que um que apresente uma relação de 85% mas que demore o dobro no tempo de entrega do pedido.

Referências

- ABRASEL. **Crise leva ao fechamento de 40% dos restaurantes de comida a quilo**. 2021. Disponível em: <<https://agenciabrasil.ebc.com.br/economia/noticia/2021-06/crise-leva-ao-fechamento-de-40-dos-restaurantes-de-comida-quilo>>. Acesso em: 09 de junho 2023. Citado na página 18.
- _____. **Delivery chega a 89% dos restaurantes brasileiros com a pandemia da Covid**. 2021. Disponível em: [Delivery chega a maior parte dos restaurantes brasileiros por conta da pandemia](#). Acesso em: 09 de junho 2023. Citado na página 18.
- _____. **Comida no iFood fica 17,5% mais cara do que no restaurante, diz Abrasel**. 2022. Disponível em: <<https://abrase.com.br/noticias/noticias/comida-no-ifood-fica-17-5-mais-cara-do-que-no-restaurante-diz-abrase/>>. Acesso em: 08 de junho 2023. Citado na página 17.
- AMARAL, L. **Operação Escola Segura: suspeito é preso e 270 contas no Twitter têm exclusões solicitadas**. 2023. Disponível em: <<https://www.cnnbrasil.com.br/nacional/operacao-escola-segura-suspeito-e-preso-e-270-contas-no-twitter-tem-exclusoes-solicitadas/>>. Acesso em: 01 de junho 2023. Citado na página 21.
- AMAZON. <https://venda.amazon.com.br/sellerblog/e-commerce-ou-marketplace-conhecas-suas-principais-diferencas>. 2021. Disponível em: <<https://venda.amazon.com.br/sellerblog/e-commerce-ou-marketplace-conhecas-suas-principais-diferencas>>. Acesso em: 12 de julho 2023. Citado na página 10.
- AMERICO, J. **Dark kitchen: o modelo de restaurantes que veio para ficar/**. VC S/A, 2021. Citado na página 19.
- ANR. **NOVA PESQUISA ANR E GALUNION APONTA ALTA NO FECHAMENTO DE OPERAÇÕES DE BARES E RESTAURANTES NA PANDEMIA**. 2020. Disponível em: [ANR aponta para alta no fechamento de bares e restaurantes](#) Acesso em: 09 de junho 2023. Citado na página 18.
- _____. **NOVA PESQUISA ANR | GALUNION | IFB: 62% DOS BARES E RESTAURANTES AINDA NÃO RETOMARAM VENDAS PRÉ-PANDEMIA; RECUPERAÇÃO PODE LEVAR PELO MENOS DOIS ANOS**. 2021. Disponível em: [Nova pesquisa da ANR, restaurantes não retomaram vendas pré-pandemia](#). Acesso em: 09 de junho 2023. Citado na página 18.
- ASANI, E.; VAHDAT-NEJAD, H.; SADRI, J. Restaurant recommender system based on sentiment analysis. **Machine Learning with Applications**, Elsevier, v. 6, p. 100114, 2021. Citado 2 vezes nas páginas 24 e 25.
- BETANCOURT, Y.; ILARRI, S. Use of text mining techniques for recommender systems. In: **ICEIS (1)**. [S.l.: s.n.], 2020. p. 780–787. Citado na página 23.

- CHEHAB, M. **Knowledge Discovery Data (KDD)**. 2020. Disponível em: <<https://medium.com/analytics-vidhya/knowledge-discovery-data-kdd-a8b41509bff9>>. Acesso em: 05 de junho 2023. Citado na página 21.
- CÔRTEZ, S. da C.; PORCARO, R. M.; LIFSCHITZ, S. **Mineração de dados-funcionalidades, técnicas e abordagens**. [S.l.]: PUC, 2002. Citado na página 23.
- DATA.AI. **S T A T E O F M O B I L E 2 0 2 1**. 2021. Disponível em: <<https://www.data.ai/en/go/state-of-mobile-2021/>>. Acesso em: 08 de junho 2023. Citado 2 vezes nas páginas 4 e 17.
- DAVIDSON, J.; LIEBALD, B.; LIU, J.; NANDY, P.; VLEET, T. V.; GARGI, U.; GUPTA, S.; HE, Y.; LAMBERT, M.; LIVINGSTON, B. et al. The youtube video recommendation system. In: **Proceedings of the fourth ACM conference on Recommender systems**. [S.l.: s.n.], 2010. p. 293–296. Citado na página 24.
- Deliverect. **WHAT IS A DARK KITCHEN (GHOST KITCHEN)? (2023)**. 2023. Disponível em: <<https://www.deliverect.com/en/blog/dark-kitchens/what-is-a-dark-kitchen>>. Acesso em: 16 de julho 2023. Citado na página 11.
- DIALANI, P. **The Future of Data Revolution will be Unstructured Data**. 2020. Disponível em: <<https://www.analyticsinsight.net/the-future-of-data-revolution-will-be-unstructured-data/>>. Acesso em: 01 de junho 2023. Citado na página 21.
- EXPERT.AI. **10 Practical Text Mining Examples to Leverage Right Now**. 2022. Disponível em: <<https://www.expert.ai/blog/10-text-mining-examples/>>. Acesso em: 01 de junho 2023. Citado na página 21.
- FLOREZ-LOPEZ, R. Strategic supplier selection in the added-value perspective: A ci approach. **Information sciences**, Elsevier, v. 177, n. 5, p. 1169–1179, 2007. Citado na página 24.
- G1. **Uber Eats deixa de funcionar no Brasil a partir desta terça**. 2022. Disponível em: <<https://g1.globo.com/tecnologia/noticia/2022/03/07/uber-eats-deixa-de-funcionar-no-brasil-a-partir-desta-terca.ghtml>>. Acesso em: 08 de junho 2023. Citado na página 17.
- GRANDI, G. **Uber Eats encerra o serviço no Brasil, saiba o que muda**. 2022. Disponível em: <<https://www.gazetadopovo.com.br/bomgourmet/negocios-e-franquias/uber-eats-encerra-brasil-o-que-muda/>>. Acesso em: 08 de junho 2023. Citado na página 17.
- HAKIM, M. P.; LIBERA, V. M. D.; ZANETTA, L. D.; STEDEFELDT, E.; ZANIN, L. M.; SOON-SINCLAIR, J. M.; WIŚNIEWSKA, M. Z.; CUNHA, D. T. da. Exploring dark kitchens in brazilian urban centres: a study of delivery-only restaurants with food delivery apps. **Food Research International**, Elsevier, p. 112969, 2023. Citado na página 19.
- HAQQI, S. **The most popular fashion brands around the world**. 2022. Disponível em: <<https://www.money.co.uk/credit-cards/most-popular-fashion-brands-2022>>. Acesso em: 08 de junho 2023. Citado na página 16.

- HARRIS, S. **Craig Cohen and Michael Adelberg take orders on-line**. 1996. Disponível em: <https://www.paloaltoonline.com/weekly/morgue/monthly/1996_Feb_28.WAITER.html>. Acesso em: 08 de junho 2023. Citado na página 17.
- HOLEWA, K. **We know what you like! Perks of recommendation systems in business**. 2020. Disponível em: <<https://www.miquido.com/blog/perks-of-recommendation-systems-in-business/>>. Acesso em: 07 de junho 2023. Citado na página 23.
- IFOOD. **Quanto cresce o uso de apps de delivery de comida no mundo?** 2022. Disponível em: <<https://news.ifood.com.br/quanto-cresce-o-uso-de-apps-de-delivery-de-comida-no-mundo/>>. Acesso em: 08 de junho 2023. Citado na página 17.
- JACKSON, E. **History Of Food Delivery And How It's Changed**. 2021. Disponível em: <<https://www.thistle.co/learn/thistle-thoughts/history-of-food-delivery-and-how-its-changed>>. Acesso em: 08 de junho 2023. Citado na página 17.
- JIVANI, A. G. et al. A comparative study of stemming algorithms. **Int. J. Comp. Tech. Appl**, v. 2, n. 6, p. 1930–1938, 2011. Citado na página 22.
- Josué de Oliveira. **O iFood comeu o mercado de delivery no Brasil**. 2022. Disponível em: <<https://tecnoblog.net/especiais/o-ifood-comeu-o-mercado-de-delivery-no-brasil/>>. Acesso em: 12 de julho 2023. Citado na página 10.
- Katie McCann. **Food delivery app boom continues following lockdown-driven growth**. 2020. Disponível em: <<https://www.touchbistro.com/blog/pros-and-cons-of-offering-food-delivery-service/>>. Acesso em: 12 de julho 2023. Citado na página 11.
- KESAVARAJ, G.; SUKUMARAN, S. A study on classification techniques in data mining. In: **IEEE. 2013 fourth international conference on computing, communications and networking technologies (ICCCNT)**. [S.l.], 2013. p. 1–7. Citado na página 23.
- KIMES, S. E.; LAQUE, P. Online, mobile, and text food ordering in the us restaurant industry. 2011. Citado na página 17.
- KURNIA, Y.; ISHARIANTO, Y.; GIAP, Y. C.; HERMAWAN, A. et al. Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the o! fish restaurant using apriori algorithm. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. [S.l.], 2019. v. 1175, n. 1, p. 012047. Citado 2 vezes nas páginas 24 e 25.
- LEMONAKI, D. **What is an Outlier? Definition and How to Find Outliers in Statistics**. 2021. Disponível em: <<https://www.freecodecamp.org/news/what-is-an-outlier-definition-and-how-to-find-outliers-in-statistics/>>. Acesso em: 05 de junho 2023. Citado na página 22.
- LIMA, L. S. de; GONÇALVES, E. C. Similaridade semântica de nomes de produtos alimentícios utilizando wordnets do português. In: **ONTOBRAS**. [S.l.: s.n.], 2022. Citado na página 25.

- LUFKIN, B. **The curious origins of online shopping**. 2020. Disponível em: <<https://www.bbc.com/worklife/article/20200722-the-curious-origins-of-online-shopping>>. Acesso em: 08 de junho 2023. Citado na página 15.
- MAGAZINE, F. **No Brasil, poucos restaurantes ficam antigos**. 2015. Disponível em: <<https://foodmagazine.com.br/food-service-noticia/no-brasil,-poucos-restaurantes-ficam-antigos>>. Acesso em: 10 de junho 2023. Citado na página 20.
- MEIRELLES, T. P.; GONÇALVES, E. C.; GOMES, D. T. Pareamento de nomes de produtos e serviços utilizando medidas de similaridade textual nos níveis alfabético, léxico e semântico. **Cadernos do IME-Série Informática**, v. 46, p. 104–117, 2021. Citado 2 vezes nas páginas 25 e 26.
- MICHAEL SCHRAGE. **On-Line Pizza Idea Is Clever but Only Half-Baked**. 1994. Disponível em: <<https://www.latimes.com/archives/la-xpm-1994-08-25-fi-31168-story.html>>. Acesso em: 12 de julho 2023. Citado na página 10.
- MIVA. **The History Of Ecommerce: How Did It All Begin?** 2020. Disponível em: <<https://blog.miva.com/the-history-of-ecommerce-how-did-it-all-begin>>. Acesso em: 12 de julho 2023. Citado na página 10.
- Mohammed Saqib. **17 Most Valuable E-Commerce Companies in the World**. 2023. Disponível em: <<https://finance.yahoo.com/news/17-most-valuable-e-commerce-171841990.html>>. Acesso em: 12 de julho 2023. Citado na página 10.
- NORMAN, J. M. **Michael Aldrich Invents Online Shopping**. 2023. Disponível em: <<https://www.historyofinformation.com/detail.php?entryid=4528>>. Acesso em: 08 de junho 2023. Citado na página 15.
- OTEMPO. **Consumidor reclama que comer fora de casa está caro e a quantidade está menor**. 2022. Disponível em: <<https://www.otempo.com.br/economia/consumidor-reclama-que-comer-fora-de-casa-esta-carro-e-a-quantidade-esta-menor>>. Acesso em: 10 de junho 2023. Citado na página 20.
- PARMAR, H. **Differences between Online store Marketplace - What to Choose?** 2021. Disponível em: <<https://www.mconnectmedia.com/blog/online-store-vs-marketplace-difference/>>. Acesso em: 08 de junho 2023. Citado na página 16.
- PARSA, H.; GREGORY, A.; TERRY, M. et al. Why do restaurants fail? part iii: An analysis of macro and micro factors. 2011. Citado 2 vezes nas páginas 11 e 20.
- PARSA, H.; SELF, J. T.; NJITE, D.; KING, T. Why restaurants fail. **Cornell Hotel and Restaurant Administration Quarterly**, Sage Publications Sage CA: Thousand Oaks, CA, v. 46, n. 3, p. 304–322, 2005. Citado na página 20.
- PORTER, D. **The Full Cost of Building an Ecommerce Website in 2023**. 2023. Disponível em: <<https://statrys.com/blog/ecommerce-website-cos>>. Acesso em: 08 de junho 2023. Citado na página 16.

PORTUGUÊS, C. do. **Stop words – o que são e como identificar?** 2022. Disponível em: <<https://www.clubedoportugues.com.br/stop-words/>>. Acesso em: 05 de junho 2023. Citado na página 22.

PRINCETON. **What is WordNet?** 2005. Disponível em: <<https://wordnet.princeton.edu/>>. Acesso em: 25 de maio 2023. Citado na página 26.

PYMNTS. **Amazon Says Third-Party Sellers Drive 60% of eCom-merce Sales.** 2023. Disponível em: <<https://www.pymnts.com/amazon/2023/amazon-independent-sellers-make-up-60percent-of-ecommerce-sales/>>. Acesso em: 08 de junho 2023. Citado na página 16.

RAKHMAWATI, N. A.; JANNAH, M. Food ingredients similarity based on conceptual and textual similarity. **Halal Research Journal**, v. 1, n. 2, p. 87–95, 2021. Citado 2 vezes nas páginas 25 e 26.

ROSA, G. S. **iFood domina o delivery no Brasil, mas restaurantes e rivais contam como vão reagir.** 2022. Disponível em: <<https://tecnoblog.net/especiais/ifood-domina-o-delivery-no-brasil-mas-restaurantes-e-rivais-contam-como-va-reagir/>>. Acesso em: 08 de junho 2023. Citado na página 17.

RUPP, I. **Restaurantes precisam atualizar forma como compram produtos e gerem seus estoques.** 2020. Disponível em: [Maneira do food-service comprar precisa se atualizar](#) . Acesso em: 10 de junho 2023. Citado na página 20.

SALESFORCE. **CRM 101: What is CRM?** 2023. Disponível em: <<https://www.salesforce.com/crm/what-is-crm/>>. Acesso em: 14 de setembro 2023. Citado na página 12.

_____. **Why Do CRM Projects Fail (And How to Fix Them).** 2023. Disponível em: <<https://www.salesforce.com/ap/hub/crm/why-do-crm-projects-fail/>>. Acesso em: 01 de maio 2023. Citado na página 12.

SCUADRA. **Dark Kitchen é a nova tendência no setor de food service.** 2020. Disponível em: <<https://www.scuadra.com.br/blog/dark-kitchen-e-a-nova-tendencia-no-setor-de-food-service/>>. Acesso em: 09 de junho 2023. Citado na página 19.

SEE-KWONG, G.; SOO-RYUE, N.; SHIUN-YI, W.; LILY, C. Outsourcing to online food delivery services: Perspective of f&b business owners. **The Journal of Internet Banking and Commerce**, Research and Reviews, v. 22, n. 2, p. 1–18, 2017. Citado na página 18.

SELF, J. T.; JONES, M. F.; BOTIEFF, M. Where restaurants fail: A longitudinal study of micro locations. **Journal of Foodservice Business Research**, Taylor & Francis, v. 18, n. 4, p. 328–340, 2015. Citado na página 11.

SHARMA, R. **KDD Process in Data Mining: What You Need To Know?** 2020. Disponível em: <<https://www.upgrad.com/blog/kdd-process-data-mining/>>. Acesso em: 05 de junho 2023. Citado na página 21.

SISCHEF. **Por que os restaurantes fecham? Saiba como evitar.** 2023. Disponível em: <<https://sischef.com/por-que-os-restaurantes-fecham/>>. Acesso em: 19 de julho 2023. Citado na página 11.

- SMALLCOMBE, M. **Structured vs Unstructured Data: 5 Key Differences**. 2023. Disponível em: <<https://www.integrate.io/blog/structured-vs-unstructured-data-key-differences>>. Acesso em: 01 de junho 2023. Citado na página 21.
- SMITH, B.; LINDEN, G. Two decades of recommender systems at amazon. com. **Ieee internet computing**, Ieee, v. 21, n. 3, p. 12–18, 2017. Citado na página 23.
- STEDMAN, C. **Text Mining (text analytics)**. 2023. Disponível em: <<https://www.techtarget.com/searchbusinessanalytics/definition/text-mining>>. Acesso em: 01 de junho 2023. Citado na página 21.
- THURAISINGHAM, B. A primer for understanding and applying data mining. **It Professional**, IEEE, v. 2, n. 1, p. 28–31, 2000. Citado na página 22.
- Tiahn Wetzler. **Food delivery app boom continues following lockdown-driven growth**. 2021. Disponível em: <<https://www.adjust.com/blog/food-delivery-apps-thrive-following-covid-growth/>>. Acesso em: 12 de julho 2023. Citado na página 10.
- VIDAL, F. **What Is A Marketplace? Our understanding of multi-seller businesses**. 2019. Disponível em: <<https://www.shoperly.com/insights/what-is-a-marketplace>>. Acesso em: 08 de junho 2023. Citado na página 16.
- VIJAYARANI, S.; ILAMATHI, M. J.; NITHYA, M. et al. Preprocessing techniques for text mining-an overview. **International Journal of Computer Science & Communication Networks**, v. 5, n. 1, p. 7–16, 2015. Citado na página 22.
- VYSHNAVI, V.; MALLIKA, C. N. Impact of corona-19 on e-commerce-a comprehensive study. **Dynamic Business Trends and Innovations in Contemporary Times**, p. 145, 2020. Citado na página 10.
- WAITER. **OUR STORY HAPPINESS THROUGH FOOD**. 2023. Disponível em: <<https://www.waiter.com/our-story>>. Acesso em: 12 de julho 2023. Citado na página 10.
- WILSON, A. **Data Mining in CRM**. 2022. Disponível em: <<https://www.rolustech.com/blog/data-mining-crm>>. Acesso em: 07 de julho 2023. Citado na página 23.
- ZHANG, S.; WU, X. Fundamentals of association rules in data mining and knowledge discovery. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 1, n. 2, p. 97–116, 2011. Citado na página 23.