



# MÉTODOS DE DETECÇÃO DE FAKE NEWS: COMPARATIVO ENTRE ABORDAGENS CROWD SIGNALS E MÉTODOS DE COMITÊ

Uriel Merola Minage e Silva

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Ciência da Computação, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ, como parte dos requisitos necessários à obtenção do grau de mestre.

Orientador: Jorge de Abreu Soares  
Coorientador: Ronaldo Ribeiro Golschmidt

Rio de Janeiro,  
Setembro de 2023

# Métodos de detecção de fake news: Comparativo entre abordagens crowd signals e métodos de comitê

Dissertação de Mestrado em Ciência da Computação, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ.

Uriel Merola Minage e Silva

Aprovada por:

JORGE DE ABREU  
SOARES:01396774700

Assinado de forma digital por JORGE  
DE ABREU SOARES:01396774700  
Dados: 2023.10.25 18:53:15 -03'00'

Presidente, Prof. Jorge de Abreu Soares, D.Sc. (Orientador)



Documento assinado digitalmente

RONALDO RIBEIRO GOLDSCHMIDT

Data: 30/10/2023 15:21:17-0300

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

Ronaldo Ribeiro Golschmidt , D.Sc. (Coorientador) (IME)



Documento assinado digitalmente

EDUARDO BEZERRA DA SILVA

Data: 28/10/2023 07:47:02-0300

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

Eduardo Bezerra da Silva, D.Sc. (Cefet/RJ)



Documento assinado digitalmente

PAULO MARCIO SOUZA FREIRE

Data: 31/10/2023 06:53:13-0300

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

Paulo Márcio Souza Freire , D.Sc. (IME)

Rio de Janeiro,  
Setembro de 2023

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

S568 Silva, Uriel Merola Minage e  
Métodos de detecção de fake news: comparativo entre  
abordagens crowd signals e métodos de comitê / Uriel Merola  
Minage e Silva. — 2023.  
48f. : il. color. , enc.

Dissertação (Mestrado) Centro Federal de Educação  
Tecnológica Celso Suckow da Fonseca, 2023.

Bibliografia : f. 45-48

Orientadora: Jorge de Abreu Soares

Coorientador: Ronaldo Ribeiro Golschmidt

1. Notícias falsas. 2.Desinformação. 3. Aprendizado do  
computador. 4. Redes sociais. I. Soares, Jorge de Abreu (Orient.).  
II. Golschmidt, Ronaldo Ribeiro. (Coorient.). III. Título.

CDD 302.23

## DEDICATÓRIA

Dedico este trabalho ao meu pai e minha irmã,  
vítimas da desinformação e da COVID-19.

## AGRADECIMENTOS

Agradeço imensamente aos meus familiares que me apoiaram durante todo o longo trajeto percorrido até aqui, pelos incentivos, paciência e confiança na minha capacidade.

Agradeço a todos os mentores que de alguma forma contribuíram para que eu chegasse aqui, em especial, ao Renato Cameira e ao Samir Azzam por me mostrarem novos caminhos e me fortalecerem para percorrê-los.

Agradeço também aos meus orientadores, Jorge Soares e Ronaldo Golschmidt, por me conduzirem e elevarem a caminhos mais altos.

## RESUMO

Métodos de detecção de fake news: Comparativo entre abordagens crowd signals e métodos de comitê

O aumento significativo da disseminação de *Fake News*, deve-se à facilidade de criação e consumo de informações nas redes sociais. Várias abordagens baseadas em aprendizado de máquina (AM) foram propostas para detectar e combater esse tipo de informação maliciosa. Entre as principais abordagens para detectar *fake news* existe uma baseada em *Hybrid Crowd Signals* (HCS). Para identificar informações falsas, esta abordagem combina sinais (i.e opiniões sobre se a informação é falsa ou não) coletados tanto dos usuários de redes sociais quanto de classificadores de AM. Embora promissora, até onde pudemos observar, a HCS emprega um método ingênuo (*Naive Bayes*) para combinar os sinais e inferir quais notícias são falsas. Assim, o presente trabalho levanta a hipótese de que métodos *ensemble* aplicados para combinar as opiniões fornecidas pelos classificadores de AM usados na HCS e as opiniões implícitas dos usuários fornecidas nas redes sociais, podem resultar em modelos de classificação com melhor desempenho. Os experimentos conduzidos neste trabalho fornecem indícios da validade da hipótese levantada.

Palavras-chave:

Notícias falsas, desinformação, aprendizado de máquina, comitês, crowd signals

Rio de Janeiro,  
Setembro de 2023

## ABSTRACT

Fake news detection methods: Comparing Crowd Signals and Ensembles Based Approaches

The significant rise of fake news dissemination is due mainly to the easy generation and consumption of information provided by social networks. Several machine learning-based approaches have been proposed to detect and combat this malicious kind of information. Among the leading approaches to detect fake news there is one based on hybrid crowd signals (HCS). To identify false information, this approach combines signals (i.e. opinions on whether the information is false or not) collected either from the users of social networks or from machine learning classifiers. Although promising, as far as we could observe, HCS employs a naive bayes classifier (i.e. a naive method) to combine the signals from the crowd and infer which pieces of news are false. Hence, the present work raises the hypothesis that ensemble methods applied to combine the opinions provided by the machine learning classifiers used in HCS and implicit opinions provided by users in social networks can lead to better classification models. The experiments conducted in this study provide evidence of the validity of the raised hypothesis.

Keywords:

Fake news, misinformation, disinformation, machine learning, ensemble, crowd signals

Rio de Janeiro,  
Setembro de 2023

## Sumário

<b>I</b>	<b>Introdução</b>	<b>14</b>
<b>II</b>	<b>Fundamentação Teórica</b>	<b>17</b>
II.1	<i>Fake News</i>	17
II.2	<i>Crowdsourcing</i>	17
II.2.1	<i>Crowd Signals</i>	18
II.2.2	<i>Crowd Híbrido</i>	18
II.3	Inteligência Artificial	18
II.4	Aprendizado de máquina	19
II.5	Aprendizado Supervisionado	19
II.6	Comitês	19
II.6.1	Método <i>Bagging</i>	20
II.6.2	Votação Majoritária	21
II.7	Método Hybrid Crowd Signals - Implicit (HCS-I)	21
II.8	Método Hybrid Crowd Signals - Full (HCS-F)	23
<b>III</b>	<b>Trabalhos Relacionados</b>	<b>24</b>
<b>IV</b>	<b>Metodologia dos Experimentos e Resultados</b>	<b>38</b>
IV.1	Formalização dos Ensembles desenvolvidos	38
IV.2	Fase 1 - Comparativo entre os ensembles desenvolvidos e o método HCS-I	38
IV.2.1	Pré-Processamento e Particionamento do Dataset na Fase 1	39
IV.2.2	Geração de Ensembles com Máquinas do HCS-F	40
IV.2.3	Geração de Ensembles Bagging da Fase 1	42
IV.2.4	Geração de Ensembles com Melhores Classificadores da Fase 1	45
IV.2.5	Comparação e Análise de Resultados da Fase 1	46
IV.3	Fase 2 - Comparativo entre os ensembles desenvolvidos e método HCS-F	47
IV.3.1	Pré-Processamento do <i>Dataset</i> na Fase 2	47
IV.3.2	Geração de Ensembles com opiniões implícitas do HCS-I e Máquinas do HCS-F	48
IV.3.3	Geração de Ensembles Bagging da Fase 2	51



IV.3.4	Geração de Ensembles com Melhores Classificadores da Fase 2	54
IV.3.5	Comparação e Análise de Resultados da Fase 2	54
<b>V</b>	<b>Conclusão</b>	<b>56</b>
V.1	Considerações Finais	56
V.2	Trabalhos futuros	57
	Referências	58

## Lista de Figuras

I.1	Exemplos de <i>fake news</i> disseminadas na pandemia de COVID-19: G1	14
II.1	Redução de variância em comitês [Zhang and Ma, 2012]	20
II.2	Ilustração do Método <i>Bagging</i> [Autoria Própria]	21
III.1	Visão estatísticas dos datasets utilizados [Freire and Goldschmidt, 2019]	35
III.2	Resultados das máquinas x HCS em cada <i>dataset</i> [Souza Freire et al., 2021]	36
IV.1	Etapas da metodologia de experimentação adotada na fase 1	39
IV.2	Acurácia dos <i>Ensembles</i> com Máquinas do HCS-F no <i>dataset</i> FakeNewsSet	40
IV.3	Acurácia dos <i>Ensembles</i> com Máquinas do HCS-F no <i>dataset</i> Gossip	41
IV.4	Acurácia dos <i>Ensembles</i> com Máquinas do HCS-F no <i>dataset</i> Gossip2	41
IV.5	Acurácia dos <i>Ensembles</i> com Máquinas do HCS-F no <i>dataset</i> PolitiFact	41
IV.6	Acurácia dos <i>Ensembles</i> com Máquinas do HCS-F no <i>dataset</i> FakeBr	42
IV.7	Acurácia dos <i>Ensembles Bagging</i> com Máquinas do HCS-F no <i>dataset</i> FakeNewsSet	43
IV.8	Acurácia dos <i>Ensembles Bagging</i> com Máquinas do HCS-F no <i>dataset</i> Gossip	44
IV.9	Acurácia dos <i>Ensembles Bagging</i> com Máquinas do HCS-F no <i>dataset</i> Gossip2	44
IV.10	Acurácia dos <i>Ensembles Bagging</i> com Máquinas do HCS-F no <i>dataset</i> PolitiFact	44
IV.11	Acurácia dos <i>Ensembles Bagging</i> com Máquinas do HCS-F no <i>dataset</i> FakeBr	45
IV.12	Acurácia dos <i>Ensembles</i> com Melhores Classificadores da Fase 1 em Cada <i>Dataset</i>	46
IV.13	Etapas da metodologia de experimentação adotada na fase 2	47
IV.14	Acurácia dos <i>Ensembles</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset</i> FakeNewsSet	48
IV.15	Acurácia dos <i>Ensembles</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset</i> Gossip	49
IV.16	Acurácia dos <i>Ensembles</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset</i> Gossip2	49
IV.17	Acurácia dos <i>Ensembles</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset</i> PolitiFact	50

IV.18	Acurácia dos <i>Ensembles</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset FakeBr</i>	50
IV.19	Acurácia dos <i>Ensembles Bagging</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset FakeNewsSet</i>	51
IV.20	Acurácia dos <i>Ensembles Bagging</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset Gossip</i>	52
IV.21	Acurácia dos <i>Ensembles Bagging</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset Gossip2</i>	52
IV.22	Acurácia dos <i>Ensembles Bagging</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset PolitiFact</i>	53
IV.23	Acurácia dos <i>Ensembles Bagging</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no <i>Dataset FakeBr</i>	53
IV.24	Acurácia dos <i>Ensembles</i> com Melhores Classificadores da Fase 2 em cada <i>Dataset</i>	54

## Lista de Tabelas

III.1	Consolidação dos trabalhos relacionados	37
IV.1	Síntese dos <i>Datasets</i> utilizados nos experimentos deste trabalho.	39
IV.2	Especificações dos <i>Ensembles</i> construídos com máquinas do HCS-F	40
IV.3	Acurácia média $\pm$ desvio-padrão dos <i>Ensembles</i> com máquinas do HCS-F no processo de validação cruzada com 10 conjuntos em cada <i>dataset</i> (Fase 1)	42
IV.4	Especificações dos <i>Ensembles bagging</i> implementados na fase 1	43
IV.5	Acurácia média $\pm$ desvio-padrão dos <i>Ensembles bagging</i> com máquinas do HCS-F no processo de validação cruzada com 10 conjuntos em cada <i>dataset</i>	45
IV.6	Especificações e resultados dos <i>Ensembles</i> com melhores classificadores - validação cruzada com 10 conjuntos em cada <i>dataset</i> (Fase 1)	46
IV.7	Especificações dos <i>Ensembles</i> construídos com modelos gerados por máquinas do HCS-F e opiniões implícitas do método HCS-I	48
IV.8	Acurácia média $\pm$ desvio-padrão dos <i>Ensembles bagging</i> no processo de validação cruzada com 10 conjuntos em cada <i>dataset</i> (Fase 2)	50
IV.9	Especificações dos <i>Ensembles bagging</i> com opiniões implícitas do HCS-I e máquinas do HCS-F	51
IV.10	Acurácia média $\pm$ desvio-padrão dos <i>Ensembles Bagging</i> com opiniões implícitas do HCS-I e Máquinas do HCS-F no processo de validação cruzada com 10 conjuntos em cada <i>dataset</i>	53
IV.11	Especificações e resultados dos <i>Ensembles</i> com melhores classificadores - validação cruzada com 10 conjuntos em cada <i>dataset</i> (Fase 2)	54

## Lista de Abreviações

**MDDN** Meios Digitais de Divulgação de Notícias

**IA** Inteligência Artificial

**AM** Aprendizado de Máquina

**HCS** *Hybrid Crowd Signals*

**RF** *Random Forest*

**SVM** *Support Vector Machines*

**KNN** *K-Nearest Neighbor*

**DT** *Decision Tree*

**NB** *Naive Bayes*

**MLP** *Multilayer Perceptron*

**LR** *Logistic Regression*

## Capítulo I Introdução

Com o crescente desenvolvimento das tecnologias de comunicação, somado às tecnologias móveis, *smartphones*, *tablets* e *notebooks*, o mundo atual está cada vez mais conectado. As informações são acessadas e compartilhadas nos meios digitais de divulgação de notícias (MDDN) (ex. jornais on-line e redes sociais) com extrema rapidez, com grande alcance e de qualquer lugar do mundo. Esse poder tecnológico e comunicativo têm ampliado a proliferação de uma categoria particular de notícia : as *Fake News*.

*Fake News* apresenta diversas definições que podem ser organizadas em dois grupos, o primeiro onde uma notícia comprovadamente falsa é divulgada intencionalmente e o segundo e mais genérico, onde uma notícia é falsa independente da intenção [Souza Freire et al., 2021]. Com a facilidade de acesso a ferramentas de criação de conteúdo, editores de imagem, texto e vídeos, a criação intencional de notícias falsas é facilitada. Muitas dessas funcionalidades estão embarcadas nas redes sociais, criando um ambiente favorável não só para a criação mas também para o compartilhamento de *fake news*.

O ambiente favorável e o comportamento viral de disseminação de *fake news* puderam ser observados durante a recente pandemia de COVID-19. Por se tratar de uma doença relativamente nova na época, de alta transmissibilidade e de potencial letalidade, diversas informações falsas sobre a forma de transmissão e tratamento foram compartilhadas, impulsionando tratamentos e medidas de proteção errôneas, impactando diretamente no diagnóstico, no tratamento adequado e possivelmente contribuindo para o aumento no número de óbitos. Exemplos de fake news divulgadas podem ser observadas na imagem abaixo.



Figura I.1: Exemplos de *fake news* disseminadas na pandemia de COVID-19: G1

Com o grande potencial de criação e disseminação, possuindo informações estruturadas e não estruturadas e mecanismos para imitar os MDDN dos grande jornais e portais verídicos, é cada vez mais difícil identificar as *fake news*. Para auxiliar nesta tarefa, as abordagens computacionais são primordiais, devido à rapidez e a capacidade de análise do grande volume de informações envolvidas.

Existem abordagens computacionais que levam em conta os dados do usuário, dados de compartilhamento, o conteúdo das notícias e utilizam a inteligência artificial (IA) e o aprendizado de máquina (AM) para auxiliar na identificação de *fake news*. As técnicas de aprendizado de máquina utilizadas são orientadas a dados, isto é, aprendem automaticamente e geram hipóteses a partir de grandes volumes de dados [Ludermir, 2021].

O desafio na utilização de algoritmos de AM é a criação de modelos que possuam altas taxas de acerto nas classificações. Para obter maior acurácia nas predições, diferentes modelos podem ser treinados e melhorados individualmente ou técnicas para classificação em conjunto, denominados comitês, podem ser empregadas para diminuir a variância e conseqüentemente aumentar a acurácia das predições [Zhang and Ma, 2012].

Existem também abordagens computacionais baseadas em *Crowd Signals*, que fazem uso das informações comportamentais dos usuários para atribuir uma reputação, obtida a partir da capacidade histórica do usuário acertar ou errar ao opinar sobre a classificação das notícias [Guo et al., 2020]. Essencialmente, as opiniões dos usuários sobre as notícias são explicitamente fornecidas através de uma funcionalidade disponibilizada pelas MDDN e são ponderadas por meio da reputação do usuário [Freire and Goldschmidt, 2019]. Apesar de possuir um potencial possivelmente robusto devido ao teor colaborativo do *Crowd signals*, esta abordagem está sensível à necessidade de explicitação da opinião pelos usuários.

Contornando a dificuldade de obtenção das opiniões explícitas dos usuários, a abordagem proposta por Souza Freire et al. [2021], denominada *Hybrid Crowd Signals* (HCS), considera que qualquer ação de divulgação (publicação ou compartilhamento) de notícia é um sinal implícito (*opinião implícita*) dado pelo usuário de que ele acredita que tal notícia é verdadeira, independentemente dessa ação ser maliciosa ou não. A abordagem HCS possui dois métodos: HCS-I e HCS-F. No HCS-I a classificação das notícias é realizada com base nas opiniões implícitas dos usuários. No HCS-F, complementar ao HCS-I, a classificação é composta, além das opiniões implícitas dos usuários, por opiniões explícitas de máquinas, ou seja, classificadores construídos a partir de algoritmos de aprendizado de máquina (AM).

Apesar dos resultados promissores do método HCS-F, o mesmo aplica inferência bayesiana ingênua como forma de combinar as opiniões explícitas das máquinas e as opiniões implícitas de usuários sobre a notícia a ser detectada (i.e., um método ingênuo). Diante disso, este trabalho levanta a seguinte hipótese: *utilizar Ensembles para combinar as opiniões explícitas das máquinas e as opi-*

*niões implícitas de usuários sobre a notícia a ser analisada pode viabilizar a construção de métodos de detecção de Fake News mais robustos que o HCS-F.*

O presente trabalho teve como objetivo realizar experimentos para avaliar se o uso de *Ensembles* como meio de integração das opiniões implícitas de usuários e as opiniões explícitas de modelos classificação gerados a partir dos algoritmos de AM poderia superar o método HCS-F. Para alcançar esse objetivo, os experimentos foram conduzidos em duas fases: A primeira fase, visando comparar tanto *Ensembles* que combinam as opiniões explícitas das máquinas do HCS-F quanto *Ensembles* compostos por modelos gerados por algoritmos de AM (diferentes das máquinas do HCS-F), com o método HCS-I (método que utiliza somente as opiniões implícitas dos usuários), objetivando encontrar evidências iniciais da plausibilidade da hipótese. A segunda fase, teve como objetivo utilizar *Ensembles* para substituir a inferência bayesiana do método HCS-F e combinar as classificações fornecidas pelas opiniões implícitas dos usuários e a opiniões explícitas fornecidas pela máquinas do método HCS-F, objetivando obter melhor desempenho do que a conjugação ingênua realizada.

Em ambas as fases os experimentos foram realizados nos mesmos *datasets* em que os métodos HCS-I e HCS-F foram originalmente avaliados. Na primeira fase foi constatado que alguns *Ensembles* obtiveram resultados superiores ao HCS-I, fortificando a possível validade da hipótese e habilitando a próxima fase de experimentos. Avançando para a segunda fase de experimentos, foi possível constatar que alguns *Ensembles* obtiveram melhor desempenho na classificação de notícias frente ao método HCS-F. Considerando os resultados obtidos, as contribuições deste trabalho foram o estudo comparativo entre a abordagem de *Crowd Signals Híbridos* e a abordagem *Ensemble* proposta, juntamente com os modelos que demonstraram um desempenho levemente superior ao método HCS-F.

A estrutura do trabalho está organizada nos seguintes capítulos: o Capítulo 2 apresentará os conceitos básicos sobre *fake news*, *crowdsourcing*, inteligência artificial, aprendizado de máquina e os métodos de comitê classificadores utilizados neste trabalho; o Capítulo 3 apresentará os trabalhos relacionados; o Capítulo 4 apresentará a metodologia dos experimentos e resultados de cada etapa; o Capítulo 5 apresentará as considerações finais e os possíveis trabalhos futuros.



## Capítulo II Fundamentação Teórica

Para o entendimento do método proposto, neste capítulo, serão apresentados os conceitos de *fake news*, *crowdsourcing*, inteligência artificial, aprendizado de máquinas e dos métodos de comitê utilizados.

### II.1 *Fake News*

Segundo Souza Freire et al. [2021] *Fake News* apresenta diversas definições que podem ser organizadas em dois grupos:

- O primeiro grupo onde uma notícia comprovadamente falsa é divulgada intencionalmente. Nesta definição a intenção proposital de compartilhamento de notícias falsas é determinante para caracterização e neste contexto a *fake news* pode ser entendida como desinformação (*disinformation*).
- O segundo grupo é mais genérico, onde uma notícia é falsa independente da intenção. Nesta definição não há conhecimento sobre a veracidade da notícia e não há intenção de desinformação, podendo ser entendido como falta de informação (*misinformation*).

Este trabalho fundamenta-se na abordagem HCS, desenvolvida por Souza Freire et al. [2021]. Portanto, adota a mesma definição de notícias falsas utilizada no estudo original, considerando-as como informações intencionalmente falsas, conforme definido no primeiro grupo.

### II.2 *Crowdsourcing*

O *crowdsourcing* pode ser caracterizado pela aplicação conjunta da inteligência humana para solucionar problemas complexos [Lofi and Maarry, 2014]. Atualmente, com auxílio da internet e redes sociais, o *crowdsourcing* é uma técnica poderosa para resolução de problemas no processamento de dados, ainda mais em casos em que os algoritmos ainda não obtiveram sucesso [Lofi and Maarry, 2014]. Apesar do grande potencial, este método é extremamente dependente da participação dos membros do *crowd* para explicitar suas opiniões (*Signals*).

### II.2.1 *Crowd Signals*

Para a emissão de um sinal explícito, os usuários fazem uso de funcionalidades embarcadas nas redes sociais para reportar uma notícia falsa ou opinar se uma notícia é falsa ou verdadeira. No contexto da classificação de *fake news* os sinais explícitos, bem como outros sinais disponíveis, podem ser utilizados para identificar um grupo de notícias potencialmente falsas [Tschischek et al., 2018]. Ainda no contexto da classificação de *fake news*, segundo Souza Freire et al. [2021] sinais implícitos podem ser inferidos a partir do comportamento dos usuários perante a notícia; tal inferência tem como princípio considerar o ato de divulgar a notícia como uma declaração, independente da intenção, de que o usuário considera a notícia verdadeira.

### II.2.2 *Crowd Híbrido*

Assim como os algoritmos apresentam dificuldades na resolução de alguns problemas, o crowdsourcing também enfrenta dificuldades relativas ao tempo, esforço e disponibilidade da força de trabalho humana nas tarefas de processamento. Uma tentativa de explorar os pontos positivos e diminuir a influência dos pontos negativos, tanto de algoritmos quanto do trabalho humano, é a utilização de um sistema híbrido de *Crowdsourcing* [Souza Freire et al., 2021], também conhecidos como *crowd* híbrido, onde algoritmos e trabalhadores humanos cooperam dinamicamente em um *workflow* combinado [Lofi and Maarry, 2014]. Segundo, Lofi and Maarry [2014] existem duas abordagens principais para os *crowds* híbridos:

- A primeira abordagem envolve a utilização de contribuições humanas para aprimorar o desempenho de algoritmos de processamento de informações. Isso pode ser alcançado fornecendo amostras de treinamento, respondendo a perguntas sobre resultados ambíguos ou fornecendo *feedback* de relevância.
- A segunda abordagem onde humanos são incluídos diretamente no processo de processamento de informações, terceirizando explicitamente algumas das tarefas ou operações necessárias dentro desse processo.

## II.3 Inteligência Artificial

O termo Inteligência Artificial (IA), exposto publicamente pela primeira vez em 1956 em uma conferência na escola superior de Dartmouth, era estudado por diversos pesquisadores em duas abordagens principais, uma de alto nível com implementações de funcionalidades inteligentes, e a outra abordagem com estudos e implementações no nível de neurônios [Brunette et al., 2009].

Existem diferentes definições de Inteligência Artificial. Para Russel and Norvig [2009], as definições giram em torno de quatro abordagens: pensamento humano, pensamento racional, ação

humana, ação racional. A IA é definida pelo estudo de agentes que recebem percepções do ambiente e executam ações [Russel and Norvig, 2009].

Diversos estudos foram realizados a partir dessas abordagens, como, por exemplo, o teste de Turing que tinha como objetivo comprovar, através de um avaliador humano, se um interlocutor era um humano ou uma máquina [Turing, 1950]. Os fundamentos da inteligência artificial permeiam pelas áreas da filosofia, psicologia, neurociência, economia, matemática e engenharia da computação, com objetivo de criar máquinas e algoritmos que simulem o pensamento e ações humanas [Russel and Norvig, 2009]. Recentemente, com o desenvolvimento de tecnologias de armazenamento, processamento e grande volume de dados gerado, a IA pode ser potencializada por meio do aprendizado de máquina.

## II.4 Aprendizado de máquina

O objetivo do aprendizado de máquina (AM) é a construção de programas que melhorem seu desempenho por meio de exemplos, onde um programa aprende uma experiência “E” com relação a uma tarefa “T” com desempenho “P” e o desempenho na tarefa melhora com a experiência [Mitchell, 1997].

O aprendizado de máquina é feito a partir de dados de exemplo. Com uma grande quantidade de dados, modelos de aprendizado podem ser desenvolvidos para o aprendizado automático. Existem três tipos principais de aprendizado de máquina: (1) aprendizado supervisionado, (2) aprendizado não supervisionado, e (3) aprendizado por reforço. Suas especificidades podem ser observadas nas subseções abaixo.

## II.5 Aprendizado Supervisionado

No aprendizado supervisionado, são fornecidos ao algoritmo de aprendizado exemplos de dados e a resposta esperada para cada exemplo; ou seja, todos os exemplos fornecidos são rotulados em uma classe. Cada exemplo é transformado em um vetor de atributos associados ao rótulo da classe apresentada. A partir desses dados, o algoritmo tem como objetivo construir um modelo que determine a classificação para novos exemplos que ainda não foram rotulados. Para rótulos discretos, obtemos uma classificação e para rótulos contínuos obtemos uma regressão Ludermir [2021].

## II.6 Comitês

Os comitês, ou sistema de tomada de decisão em conjunto, foram criados para diminuir a variância no processo de tomada de decisão, tendo por consequência o aumento da acurácia [Zhang and Ma, 2012]. Ainda segundo Zhang and Ma [2012], os comitês são fundamentados em três pilares

centrais: seleção de dados, treinamento dos modelos classificadores e combinação dos resultados dos modelos classificadores.

Segundo Zhang and Ma [2012], todo erro de classificação é composto por dois componentes controláveis: o viés e a variância, que são inversamente proporcionais, ou seja, quanto menor o viés maior a variância e vice versa. Assumindo também que a média é comumente utilizada com intuito de redução da variância nos resultados, pode-se observar na Figura 1 a composição de um comitê, em que os resultados dos diferentes modelos de aprendizado são somados e, por meio da média, é possível chegar a uma fronteira de decisão unificada com maior acurácia e baixa variância.

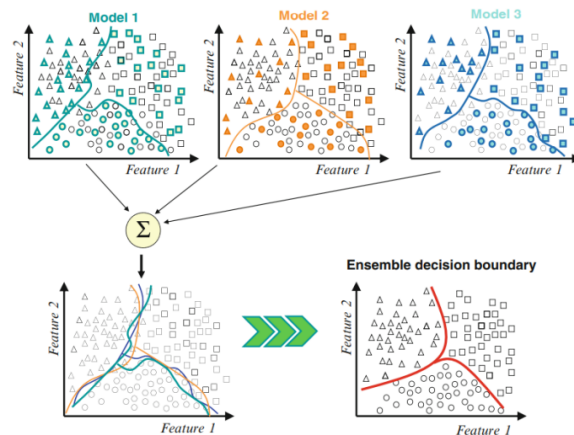


Figura II.1: Redução de variância em comitês [Zhang and Ma, 2012]

Para a a composição dos comitês, diversas abordagens podem ser adotadas. Na seção a seguir, serão apresentadas as abordagens utilizadas pelos métodos *bagging* e votação majoritária.,

### II.6.1 Método *Bagging*

O termo *bagging*, proveniente da abreviação de *bootstrap aggregating*, é um método para gerar várias versões de um preditor e utilizá-las para obter um preditor agregado [Breiman, 1996]. O método criado por Breiman [1996] consiste em gerar novas instâncias do dataset de treino, por meio da técnica de *bootstrap*, criando amostras aleatórias com reposição e utilizando-as como novos conjuntos de aprendizado. As predições realizadas em cada instância são combinadas através de votação majoritária para chegar uma previsão final.

A variabilidade introduzida pelas diferentes instâncias durante o treinamento ajuda a capturar diferentes perspectivas dos dados, tornando o modelo mais robusto a variações e ruídos nos dados de entrada e reduz a variância geral do modelo aumentando sua capacidade de generalização para novos dados. Uma ilustração do método *bagging* pode ser observado na figura II.2.

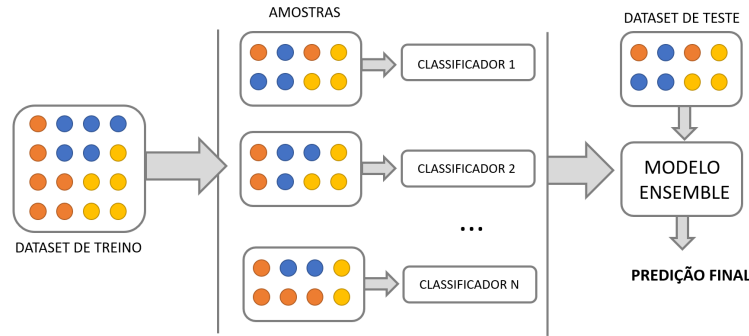


Figura II.2: Ilustração do Método *Bagging* [Autoria Própria]

### II.6.2 Votação Majoritária

Segundo Zhang and Ma [2012], a votação majoritária possui três variações: Votação unânime (onde todos os classificadores concordam), maioria simples (prevista pelo menos pela metade dos classificadores) e votação por pluralidade (que recebe o maior número de votos independente se a soma dos votos ultrapasse ou não a metade do votos). Essa técnica de combinação de classificações pode ser utilizada para conjugar as classificações de diferentes modelos a fim de obter um desempenho superior.

No contexto deste trabalho, a votação majoritária será utilizada por meio da biblioteca sklearn através do classificador denominado *VotingClassifier*<sup>1</sup>. Neste classificador é possível definir uma lista de classificadores de entrada para votação e o método de votação, dentre outros atributos.

O funcionamento da votação majoritária pode ser observado no pseudocódigo presente no algoritmo 1.

---

#### Algoritmo 1 Algoritmo de Votação Majoritária Simples

---

**Entrada:** Conjunto de classificadores  $\mathcal{H} = \{h_1, h_2, \dots, h_T\}$  e instância não rotulada  $x$

**Saída:** Rótulo de classe final  $c_{\text{final}}$  previsto para a instância  $x$

- 1: Inicialize o vetor de contagem de votos  $V$  para cada classe  $c$  com zeros
  - 2: **for**  $t = 1$  até  $T$  **do**
  - 3:     Avalie o classificador  $h_t$  em  $x$  para obter o rótulo  $c_t$
  - 4:     Incremente  $V[c_t]$  ▷ Incrementa a contagem de votos para a classe  $c_t$
  - 5: **end for**
  - 6:  $c_{\text{final}} \leftarrow \arg \max_c V$  ▷ Selecione a classe com o maior número de votos
  - 7: **retorne**  $c_{\text{final}}$  ▷ Rótulo de classe final previsto
- 

### II.7 Método Hybrid Crowd Signals - Implicit (HCS-I)

Método presente na abordagem HCS, desenvolvida por Souza Freire et al. [2021], utilizado para detecção de fake news. Este método consiste na utilização do ato de compartilhamento de uma notícia como um sinal implícito de que o usuário acredita que a notícia compartilhada é verdadeira.

<sup>1</sup><https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html>

O crowd é composto apenas pelos usuários e suas opiniões implícitas, a reputação de cada usuário é calculada com base na probabilidade de acerto e erro ao opinar sobre notícias rotuladas como fake e não fake. Para isso é criada uma matriz de opinião  $\mathcal{O}_{c_i}$  para cada usuário do crowd, onde a primeira linha é preenchida com o total de notícias compartilhadas pelo usuário que possuam o rótulo de não fake (i.e  $n_{\bar{f}\bar{f}}$ ) e pelo total de notícias compartilhadas pelo usuário que possuam o rótulo de fake (i.e.  $n_{\bar{f}f}$ ). A segunda linha da matriz  $\mathcal{O}_{c_i}$  é preenchida pela inferência da quantidade de notícias que o usuário visualizou e não compartilhou, por considerá-las como fake. Devido a indisponibilidade desta informação, a inferência é realizada com base na capacidade do usuário identificar as notícias não fake, utilizando a quantidades totais de notícias fake e não fake usando os seguintes critérios:

1. Preservar a capacidade do usuário de acertar ou errar ao opinar, ou seja,  $n_{ff}/(n_{\bar{f}\bar{f}} + n_{ff})$  deve ser aproximadamente igual a  $n_{\bar{f}\bar{f}}/(n_{\bar{f}\bar{f}} + n_{\bar{f}f})$
2. Comutar o número relativo de exemplos sinalizados nas duas classes, preservando a proporcionalidade apresentada na primeira linha de  $\mathcal{O}_{c_i}$ , ou seja,  $n_{\bar{f}\bar{f}}$  deve ser dado por  $(n_{\bar{f}f}/n_f) \times n_{\bar{f}}$ .

Uma vez que a matriz  $\mathcal{O}_{c_i}$  esteja preenchida para todos os usuários do crowd, são inferidas as probabilidades de cada usuário acertar sua opinião (i.e.  $\theta_{c_i,f}$ ) conforme indicado abaixo:

- $\theta_{c_i,f} = n_{ff}/(n_{\bar{f}\bar{f}} + n_{ff})$
- $\theta_{c_i,\bar{f}} = n_{\bar{f}\bar{f}}/(n_{\bar{f}\bar{f}} + n_{\bar{f}f})$

Uma vez que as probabilidades estejam inferidas as reputações  $rep_i$  são calculadas e registradas para cada usuário do crowd. Cada  $rep_i$  representa a matriz de probabilidades de acerto e erro, genericamente definidas abaixo.

$$rep_i = \begin{vmatrix} \theta_{c_i,\bar{f}} & 1 - \theta_{c_i,f} \\ 1 - \theta_{c_i,\bar{f}} & \theta_{c_i,f} \end{vmatrix}$$

onde:

- $\theta_{c_i,\bar{f}} = P(Y_{c_i}(n^R) = \bar{f} \mid Y^*(n^R) = \bar{f})$
- $1 - \theta_{c_i,\bar{f}} = P(Y_{c_i}(n^R) = f \mid Y^*(n^R) = \bar{f})$
- $\theta_{c_i,f} = P(Y_{c_i}(n^R) = f \mid Y^*(n^R) = f)$
- $1 - \theta_{c_i,f} = P(Y_{c_i}(n^R) = \bar{f} \mid Y^*(n^R) = f)$

Por fim, a classificação de cada notícia é realizada a partir da combinação das reputações  $rep_i$  dos usuários e suas respectivas opiniões, aplicadas em uma inferência Bayesiana que objetiva concluir se a notícia é fake ou não fake, através das equações listadas abaixo:

$$P(Y^*(n^D) = f) = \omega \cdot \prod_{c_i \in \alpha(n^D)} (1 - \theta_{c_i, f}) \quad (\text{II.1})$$

$$P(Y^*(n^D) = \bar{f}) = (1 - \omega) \cdot \prod_{c_i \in \alpha(n^D)} \theta_{c_i, \bar{f}} \quad (\text{II.2})$$

Ambas as equações se beneficiam dos acertos e dos erros de cada usuário, a partir da sua probabilidade de acertar o errar ao opinar sobre uma nova notícia e a classificação final é dada pela maior probabilidade aferida.

## II.8 Método Hybrid Crowd Signals - Full (HCS-F)

Método complementar ao HCS-I, também desenvolvido por Souza Freire et al. [2021], que utiliza além das opiniões implícitas de usuários, opiniões explícitas de modelos de classificação adquiridas através de algoritmos de aprendizado de máquina. Neste método as classificações obtidas pelos algoritmos RF, XGBOOST, SVM, DMText e FNE são incorporadas, criando um crowd híbrido, onde as reputações são calculadas tanto para usuários quanto para máquinas. A reputação das máquinas é aferida de forma semelhante às reputações dos usuários, diferenciando-se nas condições de construção da matriz de opinião  $\mathcal{O}_{c_i}$ , onde são aplicadas as condições abaixo:

- Se  $Y^*(n^R) = \bar{f}$  e  $Y_{c_i}(n^R) = \bar{f}$  então  $n_{\bar{f}\bar{f}} = n_{\bar{f}\bar{f}} + 1$ ;
- Se  $Y^*(n^R) = \bar{f}$  e  $Y_{c_i}(n^R) = f$  então  $n_{f\bar{f}} = n_{f\bar{f}} + 1$ ;
- Se  $Y^*(n^R) = f$  e  $Y_{c_i}(n^R) = \bar{f}$  então  $n_{\bar{f}f} = n_{\bar{f}f} + 1$ ;
- Se  $Y^*(n^R) = f$  e  $Y_{c_i}(n^R) = f$  então  $n_{ff} = n_{ff} + 1$ .

Na sequência, as probabilidades de acerto e erros são calculadas, e a reputação  $rep_i$  de cada membro do crowd, seja usuário ou máquina, são registradas na matriz de probabilidades de acerto e erro. Por fim, a classificação é realizada análoga ao HCS-I, onde uma inferência Bayesiana combina as reputações  $rep_i$  dos usuários ou máquinas e suas respectivas opiniões através das equações II.1 e II.2 e a classificação final é dada pela maior probabilidade aferida.

### Capítulo III Trabalhos Relacionados

Neste capítulo, serão apresentados os trabalhos relacionados a detecção automática de fake news que utilizaram algoritmos de aprendizado de máquinas e modelos de comitê. Será apresentado também o trabalho base, referente à detecção de fake news utilizando métodos de *crowd signals* implícitos e explícitos.

A pesquisa foi realizada na base de dados Scopus, por se tratar de um dos principais indexadores de artigos científicos da área de computação na atualidade. Inicialmente a busca foi realizada com a string `("fake news"OR "misinformation"OR "disinformation") AND ("machine learning") AND ("ensemble"AND "crowd signals")`; porém, não foram obtidos resultados, apontando assim uma possível análise de desempenho única entre métodos de comitê e crowd signals para a tarefa de detecção de fake news.

Na sequência, a string de busca foi alterada para `("fake news"OR "misinformation"OR "disinformation") AND ("machine learning") AND ("ensemble"OR "crowd signals")`, ainda realizada na base Scopus, obtendo 122 artigos relacionados. Após a revisão dos 122 artigos, foram selecionados 37 artigos que atuavam na detecção de fake news utilizando métodos de comitê que obtiveram melhor desempenho do que os algoritmos de aprendizado de máquina e um artigo base para o presente estudo, referente à detecção de fake news a partir de *crowd signal* híbridos. Dos 38 artigos selecionados, um foi descartado por não ter como objetivo central a detecção de fake news, e seis não estavam disponíveis de forma gratuita. Os 31 artigos selecionados que têm relação direta com a presente pesquisa foram revisados, resumidos e serão apresentados a seguir.

*T1 - A Feature-Based Approach for Sentiment Quantification Using Machine Learning [Ayyub et al., 2022]*: O artigo apresenta um modelo baseado em comitê de algoritmos de *deep learning* para a detecção de fake news relacionadas ao COVID-19 que foram divulgadas no Twitter. Os modelos CT-BERT, Bertweet e roberta foram utilizados para tarefa de detecção de *tweets* falso e tiveram seus desempenhos comparados com os algoritmos SVM, RF, LR, KNN, NB, Decision Tree, Gradiente *Boosting* e AdaBoost, e outros algoritmos de *deep learning*. O dataset utilizado consiste em 5300 publicações classificadas como fake e 5400 publicações classificadas como reais, divididos em 3060 amostras classificadas como fake e 3360 classificadas como real para o treinamento, 1120 amostras classificadas como fake e 1020 classificadas como real para validação, 1120 classificadas como fake e 1020 classificadas como real para teste. A melhor performance foi obtida pela combinação dos



algoritmos TC-BERT, Bertweet e roberta, que foram combinados através de votação, alcançando F1-score de 99%.

*T2 - Combating the infodemic: COVID-19 induced fake news recognition in social media networks* [Biradar et al., 2022]: O artigo apresenta um método de detecção de fake news relacionadas ao Covid-19, combinando atributos-chave extraídos a partir de algoritmos de análise de sentimentos, tais como BERT, XLNet e ELMo, e utilizando modelos de aprendizado de máquina e comitês para tarefa de classificação. O dataset utilizado consiste em 5600 notícias reais e 5100 notícias falsas. Foram realizadas tarefas de pré-processamento para limpeza de texto e aplicada as técnicas de extração de atributos a partir do texto das notícias. Foram criados quatro modelos de classificação, onde o Modelo 1 é formado por uma rede neural profunda com fusão precoce, o Modelo 2 por um comitê de rede neural recorrente, o Modelo 3 utiliza um comitê por votação, e o Modelo 4 consiste em um operador bit-wise multinível. Os resultados obtidos demonstraram que o modelo de comitê proposto pelo Modelo 3, que é constituído dos classificadores BERT, ULMFit com Regressão Logística, obteve resultado superior aos demais modelos executados, obtendo 98% de acurácia na classificação de fake news.

*T3 - FNU-BiCNN: Fake News and Fake URL Detection using Bi-CNN* [Sandrilla and Devi, 2022]: Este trabalho apresenta o modelo FN-BICNN para identificação de fake news e URLs falsas. O dataset utilizado possui 1574 notícias, nas quais 598 são reais e 976 falsas. Na etapa de pré-processamento, foi realizada a remoção de stopwords e lematização seguida da extração de features a partir das técnicas de *Term Frequency-Inverse Document Frequency* (TF-IDF), word2vec e POS tagging. O treinamento dos dados foi realizado a partir da mesclagem dos modelos LSTM, ARIMA e CNN. Após o treinamento, foram utilizados os algoritmos de aprendizado DT, NB, RF, SVM e k-NN para composição de um comitê de votação para obtenção da melhor acurácia na classificação; com 99,93%, o comitê por votação foi o melhor classificador encontrado.

*T4 - HAN, image captioning, and forensics ensemble multimodal fake news detection* [Meel and Vishwakarma, 2021]: Neste trabalho, é proposto um modelo de detecção de fake news multimodal, em que são utilizadas diferentes técnicas para analisar ocorrências com (1) somente textos, (2) textos e imagens, (3) imagens com texto, e (4) puramente imagens. Na primeira abordagem, foi realizada na etapa de pré-processamento a limpeza de texto, lematização, *stemming*, tokenização e extração de features baseadas na representação de *tokens* de textos a partir da técnica GloVe. Após o pré-processamento, uma rede neural denominada "Hierarchical Attention Network"(HAN) é utilizada para classificação. Constituída de quatro partes, a primeira para codificação de palavras, seguida de um sistema de atenção para obtenção de palavras relevantes, na terceira parte é realizada a codificação de sentenças, também seguida por uma parte de atenção para obtenção de sentenças mais úteis. O vetor criado com as informações é utilizado para classificação a partir da função

*softmax*.

Para a classificação de imagens, quando a imagem possui algum texto incorporado, foi utilizado a ferramenta Microsoft captionBot para criação da legenda. Tanto o texto quanto a legenda são convertidos em vetores de palavras a partir da técnica GloVe e são combinados de acordo com a semelhança semântica, caso tenham a mesma semântica e mesmo assunto são classificados como reais; caso contrário, classificados como *fakes*.

Quando a imagem for somente ilustrativa, sem conteúdo textual, é feita uma análise forense do nível de erro da imagem baseada no ruído identificado nos pixels após a edição e salvamento, sendo assim capaz de identificar uma e classificar uma imagem como manipulada ou real.

Por fim, as classificações realizadas por cada modelo são agregadas a partir da votação máxima, obtendo o modelo com melhor desempenho em relação aos modelos isolados. O modelo proposto apresentou acurácia de 95%.

*T5 - Fake News Detection - A Comparative Study of Advanced Ensemble Approaches* [Ganesh et al., 2021]: Este artigo criou um modelo a partir de um comitê de votação composto pelos algoritmos Linear SVM (LSVM), Naive Bayes (NB) e *Decision Tree* (DT) e utilizou este modelo como base para utilização de métodos *bagging* e *boosting*. Foram utilizados dois *datasets* relacionado a fake news, o primeiro dataset possuía aproximadamente 4050 notícias enquanto o segundo dataset possuía aproximadamente 20800 notícias. Na etapa de pré-processamento foi realizada a limpeza de texto através da remoção de *stopwords*, *stemming*, remoção de dados faltantes e padronização do texto em caixa baixa. Após a limpeza, foi realizada a extração de novas features a partir da contagem de frequência de palavras utilizando a técnica CountVectorizer. Após a finalização do pré-processamento, o dataset foi dividido em 70% das amostras para treinamento e 30% para teste. Os modelos LSVM, NB, DT foram treinados e combinados através de votação. Após a junção, o modelo criado foi utilizado como base nos algoritmos Bagging, Adaboosting e *Gradient Boosting*. O melhor desempenho foi observado no método Bagging com 93% de acurácia.

*T6 - A Deep Ensemble Model for News Classification on Social Media* [Rahman et al., 2021]: Neste trabalho, é apresentando um modelo ensemble criado a partir de algoritmos de *deep learning* denominado DEM. O modelo é composto pelos modelos: rede neural convolucional (CNN), rede neural profunda (DNN) e rede neural recorrente (RNN). O dataset utilizado possui cerca de 17994 notícias, onde 7607 são classificadas como falsas e 10387 classificadas como reais, 80% foi destinado para treinamento dos modelos e 20% para a realização de testes. O pré-processamento realizado consiste na remoção de stopwords, tokenização e stemming. A extração de features foi realizada a partir das técnicas de vetorização GloVe e FastText *Embedding*. Após o pré-processamento, foi criado o modelo ensemble DEM, que é composto pelos algoritmos de deep learning CNN, DNN e RNN/LSTM, agregados a partida da votação máxima. Na sequência, também foram utilizados

algoritmos de aprendizado de máquina SVM, LR, DT, NB, K-NN para comparação dos resultados. Por fim o comitê gerado pelos algoritmos de *deep learning* superou o desempenho dos modelos de aprendizado de máquinas e aprendizado profundo executados isoladamente, obtendo acurácia de 91,6%, utilizando a técnica de vetorização FastText e 90,04% ao utilizar a técnica de vetorização GloVe.

*T7 - Deep Ensemble Approach for COVID-19 Fake News Detection from Social Media* [Priya and Kumar, 2021]: O artigo propõe um modelo de comitê com abordagem de aprendizado profunda, constituído pelos algoritmos, SVM, CNN e DNN. O dataset utilizado possui 5300 notícias classificadas como falsas e 5400 classificadas como reais. Das 5300 notícias *fakes*, 3060 foram utilizadas para treinamento, 1120 para teste e 1120 para validação. Das 5400 notícias reais, 3360 foram utilizadas para treinamento, 1020 para teste e 1020 para validação. Foi utilizada a técnica TF-IDF para extração do vetor de frequência das palavras para geração de features para os algoritmos SVM e DNN, enquanto as features extraídas para o modelo CNN foram obtidas através da técnica de codificação *One-Hot*. O comitê foi agregado a partir da função de ativação softmax, que gerou a probabilidade de classificação para cada modelo utilizado. Também foram utilizados isoladamente os algoritmos de aprendizado SVM, RF, LR, K-NN, NB, GB, DT e AdaBoost para realização de comparativo de desempenho em relação ao comitê gerado. Dentre os modelos utilizados, o modelo ensemble obteve melhor desempenho utilizando a extração de features TF-IDF utilizando o nível de carácter, com F1-Score de 97%.

*T8 - Fake News Detection on English News Article's Title* [Castillo et al., 2021]: Este artigo apresenta um modelo de detecção de notícias falsas em inglês utilizando algoritmos de aprendizado de máquina e métodos ensemble. O dataset utilizado consiste em 20800 notícias classificadas possuindo os atributos ID, Título, autor, texto e classificação. O pré-processamento realizado consiste na remoção de notícias em línguas diferentes do inglês por meio da biblioteca “langdetec”, remoção de linhas nulas, remoção de notícias com menos de 20 palavras no atributo texto e com atributo título com menos de duas palavras. Após o pré-processamento, as 19500 notícias restantes foram divididas em 80% para treinamento dos modelos e 20% para a realização de testes. Na etapa de extração de features, foram extraídos atributos baseados na diversidade e densidade lexical, atributos baseados nos sentimentos foram extraídos a partir da ferramenta VADER disponível na biblioteca NLTK, atributos baseados em análises sintáticas, atributos a partir da contabilização de classes gramaticais e atributos psicolinguísticos. O modelo proposto consiste na criação de comitês utilizando o método Bagging com os algoritmos SVM, NB e XGboost como algoritmos base. Os comitês criados foram testados com diferentes configurações de *features* extraídas. Dentre os comitês gerados e dos modelos de aprendizados utilizados, o comitê Bagging, utilizando o XGboost como base, obteve maior desempenho de predição com acurácia de 79%.

*T9 - Legitimacy: An Ensemble Learning Model for Credibility Based Fake News Detection* [Ramkissoon and Goodridge, 2021]: O presente trabalho apresenta o modelo Legitimacy que foi criado para detecção de fake news a partir de um modelo de comitê formado por uma árvore de decisão utilizada como base para método *boosting* e uma rede neural profunda. O dataset utilizado possui 14000 postagens do twitter relacionadas a catástrofes naturais; 65% das amostras foram direcionadas para o treinamento dos modelos e 35% destinadas para testes. Os atributos extraídos são: texto da postagem, testemunha ocular, fonte, data, idioma, quantidade de usuário que adicionaram em suas listas, localização, fuso horário, idioma do usuário, contagem de amigos e nome de usuário. Estes atributos foram separados em duas naturezas: demográfica e comportamento social. Após o pré-processamento, limpeza de texto e normalização, os atributos referentes ao comportamento social foram selecionados para treinamento do modelo *Gradient Boost* com DT como algoritmo base e os atributos demográficos foram treinados com a DNN. Posteriormente, os resultados gerados pelos preditores foram utilizados para treinamento do modelo LR e obtenção da classificação final, que obteve acurácia de 95,3%.

*T10 - Detection of Online Fake News Using Blending Ensemble Learning* [Hansrajh et al., 2021]: O estudo propõe um modelo de comitê gerado a partir da técnica de *blending*, que consiste na utilização da predição realizada pelos algoritmos-base no dataset de teste, como entrada para o treinamento de um novo modelo. Os *datasets* utilizados são: Liar, com 12836 pequenas sentenças com 6 possíveis classificações, no qual foram aglutinadas e filtradas para compor apenas as classificações fake ou real. O segundo dataset utilizado é o ISOT com 44898 notícias sendo elas 23481 falsas e 21417 verdadeiras. Ambos os dataset possuem texto, título, assunto e data como atributos. Na etapa de pré-processamento, foram realizadas a limpeza de dados irrelevantes, tratamento de dados faltantes, conversão de texto para caixa baixa, remoção de pontuação, remoção de *stopwords* e *stemming*. A extração de features foi realizada pelos métodos TF, TF-IDF e n-grams. O treinamento dos modelos base para geração da entrada do comitê utilizando *blending* foi feito com os algoritmos LR, SVM, análise discriminante, gradiente descendente e *Ridge Regression*. O comitê gerado obteve maior desempenho em relação aos resultados dos algoritmos isoladamente, com 60,81% de acurácia no dataset LIAR e 98,41% de acurácia no dataset ISOT.

*T11 - Study and analysis of unreliable news based on content acquired using ensemble learning (prevalence of fake news on social media)* [Khan and Alhazmi, 2020]: Neste trabalho, é desenvolvida uma abordagem para classificação de fake news a partir do texto das notícias utilizando os algoritmos SVM, RF, NB, DT como algoritmos base para os métodos Bagging e AdaBoost. O dataset utilizado consiste em 20008 notícias publicadas na internet com os atributos: ID, Texto, título, autor e classificação. A extração de features foi realizada a partir do método de vetorização Doc2Vec. Após o pré-processamento e extração de features, os modelos base foram utilizados para criação de

comitês *bagging* e AdaBoost. Os comitês foram treinados e testados utilizando a técnica validação cruzada com 10 partições. Ao final, os algoritmos LSVM e RF utilizados como base para o método AdaBoost, obtiveram os melhores desempenho com 90,7% e 80,17% de acurácia respectivamente.

*T12 - An ensemble technique to detect fabricated news article using machine learning and natural language processing techniques* [Sangamnerkar et al., 2020]: O artigo realiza um comparativo entre diversos modelos de aprendizado de máquina para a classificação de notícias falsas utilizando o dataset “Fake News” com 20800 artigos de diversas fontes. Dessas, 5200 foram separadas para a realização de testes. A etapa de pré-processamento realizou a limpeza de dados, remoção de *stopwords*, de pontuação e de caracteres especiais. A extração de features foi realizada com a técnica de vetorização *doc2vec*. Após a extração de novas features, os algoritmos SVM, NB, DT, LR e Bagging foram treinados isoladamente. Após o treinamento e teste, foram realizadas diversas combinações a partir de votação pesada, onde o comitê formado pelos algoritmos DT, LR e Bagging obtiveram o melhor desempenho com acurácia de 88,08%.

*T13 - Multiclass Fake News Detection using Ensemble Machine Learning* [Kaliyar et al., 2019]: O trabalho apresenta um modelo de comitê baseado no algoritmo *Gradient Boosting* para tarefa de detecção de fake news em um dataset com quatro classificações possíveis que envolvem a correlação entre o título e o texto das notícias. O dataset consiste em 49972 instâncias, possuindo os atributos, título, texto e posicionamento, com os valores: concorda, discorda, discute e não relacionado. A etapa de extração de features contempla as técnicas de vetorização TF-IDF, similaridade dos cossenos, sobreposição de palavras, polaridade e refutação. Foram utilizados os modelos RF, NB, DT, LR, LSVM para comparação de desempenho com o modelo de comitê proposto baseado no algoritmo *gradient boosting*. O modelo proposto obteve melhor desempenho com acurácia de 86%.

*T14 - Behind the cues: A benchmarking study for fake news detection* [Gravanis et al., 2019]: Este trabalho apresenta um estudo de *benchmark* para detecção de fake news. Nele, são realizadas quatro abordagens diferentes: a primeira para propor uma combinação de features para detecção de decepção em narrativas escritas. A segunda abordagem realiza um estudo de *benchmark* extensivo sobre os algoritmos de classificação, contemplando os algoritmos SVM, KNN, DT, AdaBoost e *Bagging*. A terceira abordagem tem como objetivo descrever regras para a criação de um dataset não tendencioso para a detecção de fake news. A partir dos estudos realizados, o dataset UNBiase foi criado a partir de notícias checadas pelas entidades Snopes, TruthOrFiction e Polifact, seguindo as regras levantadas no estudo. A extração de features foi realizada a partir da técnica de vetorização *word2vec*. Em seguida, os algoritmos abordados foram treinados nos dataset gerado pelo estudo (UNBiased) e gerados pelas entidades BuzzFeed, Polifact, McIntire e Kaggle-EXT. O comitê AdaBoost obteve maior desempenho em todos os *datasets* testados, atingindo acurácia de 95%.

*T15 - Early multi-class ensemble-based fake news detection using content features* [Rezaei et al.,

2022]: O artigo propõe um modelo de detecção de fake news utilizando *Stacking*, com cinco classificadores na camada de base: Random Forest, SVM, Decision Tree, LGBM e XGBoost. O algoritmo utilizado na meta aprendizagem da segunda camada foi o AdaBoost. O modelo foi treinado e testado em um conjunto de dados criado a partir de 24.517 notícias provenientes dos sites de checagem: PoliFact, Snopes e Fact-checking, contendo notícias com rótulos binários e multi-classe. A técnica de extração de *features* utilizada foi a extração de características linguísticas, incluindo características textuais, de sentimento, semânticas e de legibilidade. As *features* utilizadas incluem o número de caracteres, o número de palavras, o número de frases, a média de caracteres por palavra e a média de palavras por frase. Vários modelos foram utilizados para o comparativo de desempenho, dentre eles: SVM, Decision Tree, Random Forest, XGBoost e LGBM. O modelo proposto obteve o melhor desempenho em comparação com os demais modelos, tanto na classificação binária, obtendo 96% de acurácia, quanto na classificação multi-classe, obtendo 83% de acurácia.

*T16 - An Enhanced Approach for Fake News Detection using Ensemble Techniques* [Gethsia and Juliet, 2023]: O artigo propõe um modelo aprimorado para detecção de *fake news* usando técnicas de aprendizado de máquina em conjunto com um classificador de votação. Foram utilizados vários *datasets* de diferentes campos, incluindo política, notícias mundiais e notícias governamentais disponibilizados no portal Kaggle. Os *datasets* estão em inglês e contêm uma quantidade variável de notícias verdadeiras e falsas. A técnica de extração de *features* foi realizada pelas técnicas TF-IDF e contagem de vetores. Os modelos utilizados para comparação de desempenho incluem NB, RF, SVM e LR. O modelo ensemble por votação contendo os modelos NB, LR, SVM e DT obteve o melhor desempenho com uma acurácia de 99,59

*T17 - An automated multi-web platform voting framework to predict misleading information proliferated during COVID-19 outbreak using ensemble method* [Varshney and Vishwakarma, 2023]: Neste trabalho, um modelo de detecção de *fake news* baseado e votação utiliza 10.700 notícias rotuladas como fake ou real do *dataset* "Constraint-2021". Cada notícia analisada é pré-processada (limpeza de texto e tradução) e enriquecida com informações provenientes das plataformas web Google e Youtube, através de uma *query* que é elaborada e incorporada junto ao termo "*Fake News*", esta *query* objetiva extrair fatos relevantes das plataformas *web* e acrescentar *features* relacionadas ao conteúdo, linguagem, similaridade e sentimento de cada notícia. Os experimentos foram realizados separadamente para cada plataforma, os modelos formados pelos algoritmos DT, LR, SVM, NB e MLP foram utilizados para comparação de desempenho. No comparativo de notícias enriquecidas pelo Google, o modelo proposto baseado em votação, formado pelos algoritmos LR, CART e LSVM obtiverem maior desempenho, apresentado 98,88% de *F1-Score*, em contra partida as notícias enriquecidas pelo YouTube não obtiveram melhor desempenho em relação aos demais algoritmos testados e apresentou o maior desempenho no modelo baseado em RF, alcançando ape-

nas 86% de *F1-Score*.

*T18 - Data Exploration and Classification of News Article Reliability: Deep Learning Study* [Zhan et al., 2022]: O trabalho utilizou técnicas de NLP e AM para construir um modelo que identifique a confiabilidade de notícias online. O dataset utilizado foi o ReCOVeRY, contendo 2029 notícias em inglês relacionadas ao COVID-19 de janeiro a maio de 2020. O dataset foi pré-processado utilizando lematização, remoção de *stop words* e remoção de artigos duplicados. Após a limpeza textual foram geradas *features* textuais e um *score* de sentimentos, também foram gerado *features* a partir da tokenização e *embendding*. Primeiramente as *features* geradas por tokenização e *embendding* foram utilizadas para treinamento dos modelos NB, KNN, LR, LSTM, BiLSTM, GRU, BiGRU e CNN. A predição do modelo de melhor desempenho na primeira fase (BiGRU), foi utilizada junto das *features* textuais e sentimentais para o treinamento de um novo modelo baseado em XGboost. O modelo proposto alcançou o maior desempenho quando comparado aos demais modelos, apresentando 0,906 de AUC.

*T19 - A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection* [Alghamdi et al., 2022]: O artigo apresenta um estudo comparativo de técnicas de aprendizado de máquina e *deep learning* para detecção de fake news. Foram utilizados diversos modelos, incluindo algoritmos clássicos de ML (LR, SVM, DT, NB, RF e XGboost), modelos avançados de ML (CNN, BiLSTM, BiGRU, CNN-BiLSTM e CNN-BiGRU) e modelos baseados em *transformers* (BERTbase e RoBERTabase). Os *datasets* utilizados foram LIAR, PolitiFact, GossipCop e COVID-19, todos em inglês, com quantidades variadas de notícias verdadeiras e falsas. A técnica de extração de *features* utilizada foi TF-IDF. Os modelos foram comparados em termos de acurácia, *F1-score* e outras métricas. O modelo que obteve o melhor desempenho foi um modelo híbrido CNN-LSTM/CNN-GRU com *embeddings* GloVe, com uma acurácia de 94,7%.

*T20 - Tackling the infodemic during a pandemic: A comparative study on algorithms to deal with thematically heterogeneous fake news* [Vasist and Sebastian, 2022]: O artigo propõe um modelo de detecção de *fake news* que utiliza uma combinação de técnicas de aprendizado de máquina, *deep learning* e *ensemble*. Foram utilizados dois *datasets*, um com notícias tematicamente heterogêneas que incluíam notícias também relacionadas ao COVID-19 e outro composto por notícias relacionadas somente ao COVID-19. Ambos os *datasets* contêm notícias em inglês e possuem uma quantidade similar de notícias verdadeiras e falsas. A técnica de extração de *features* utilizada foi a CountVectorizer e a TF-IDF. Foram utilizados vários modelos para comparar o desempenho, incluindo MMNB, *Passive Agressive Classifier*, *Multinomial classifier*, *Max Voting* contendo SVM, DT e LR, RF, Adaboost XGboost e RNN-LSTM. O modelo que obteve o melhor desempenho utilizou o classificador *Max Voting* contendo SVM, DT e LR, com uma acurácia de 98,4% e 98,3% no dataset heterogêneo utilizando as *features* geradas pelo CountVectorizer e TF-IDF respectivamente

e 95,5% e 95,1% no dataset possuindo apenas notícias relacionadas ao COVID-19 utilizando as *features* geradas pelo CountVectorizer e TF-IDF respectivamente.

*T21 - Multimodal Fake-News Recognition Using Ensemble of Deep Learners* [Al Obaid et al., 2022]: O artigo propõe um modelo de reconhecimento de fake news baseado em um ensemble de *deep learners*, que utiliza uma função de perda que enfoca diferentes aspectos da notícia por meio de um mecanismo de atenção. Os *datasets* utilizados para treinamento e teste foram o PoliFact e o Gossip, ambos em inglês, com conteúdo textual, imagens e comentários de usuários sobre as notícias. A técnica de extração de *features* utilizada para a parte textual é realizada em três etapas, a primeira denominada *sentence encoder*, onde as palavras são convertidas em vetores a partir das técnicas de *word embedding* utilizando BERT, word2vec, Glove e FastText, em seguida a etapa *post encoder* para representação vetorial de cada frase, por fim o módulo de atenção sinaliza pesos para as partes mais importantes das notícias. Para a extração de *features* de imagens, o modelo pré-treinado VGG19 é utilizado. As *features* utilizadas incluem informações do título, corpo da notícia, imagem e contexto social. Os modelos utilizados para comparação de desempenho foram o SpotFake-Plus e Fake-Bert. O modelo proposto, denominado DEFD, obteve a maior acurácia, com 85,43% no *dataset* Gossip e 85,58% no *dataset* PolitiFact.

*T22 - Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique* [Ali et al., 2022]: O artigo apresenta um modelo *ensemble* de *deep learn*, que utiliza uma técnica de aprendizado profundo sequencial para detectar *fake news*. Foram utilizados dois *datasets* com notícias em inglês, o LIAR e o ISOT. As *features* foram extraídas do conteúdo textual das notícias, enriquecidas com a técnica n-grams e representadas utilizando TF-IDF. Na primeira fase de construção do modelo, classificadores binários foram utilizados para extrair características ocultas das notícias, na sequência o algoritmo MLP foi utilizado para realização da classificação final. O modelo proposto foi comparado com o desempenho de modelos gerados a partir dos algoritmos CNN, RF, LSVM e RF utilizados nos estudos relacionados, alcançando uma precisão de detecção de 100% no conjunto de dados ISOT e 51% no dataset LIAR.

*T23 - Fake News Detection Using Machine Learning Algorithms* [Imbwaga et al., 2022]: O artigo propõe um sistema de detecção de *fake news* usando algoritmos de aprendizado de máquina supervisionado. O dataset utilizado para treinamento e teste do modelo foi obtido do Kaggle, consistindo de notícias verdadeiras e falsas em inglês, com 21418 e 23503 entradas, respectivamente. A técnica de extração de *features* utilizada foi a TF-IDF. As *features* utilizadas foram o título e o texto das notícias. Os modelos utilizados para comparar o desempenho foram LR, XGBoost, DT e RF. O modelo gerado pelo XGboost obteve maior desempenho, atingindo 99,5% de acurácia.

*T24 - A heuristic-driven uncertainty based ensemble framework for fake news detection in tweets and news articles* [Das et al., 2022]: Este artigo apresenta um *framework* de detecção de *fake*



*news* em tweets e notícias que utiliza um modelo *ensemble* de *deep learning* e uma heurística para classificação. Foram utilizados dois *datasets* para treinamento e teste do modelo, ambos em inglês: o dataset COVID-19 fake news e o dataset FakeNewsNet, composto pelos datasets PolitiFac e Gossip. As *features* são extraídas a partir da tokenização a partir da combinação das técnicas BERT, RoBERTa, XML-RoBERTa, DeBERTa, ERNIE 2.0 e ELECTRA. Além dos termos vetorizados são adicionados *features* estatísticas adquiridas a partir dos metadados das notícias. O *framework* proposto também quantificou a incerteza preditiva confiável juntamente com o nível de confiança apropriado na saída de classe para a tarefa de classificação. O *framework* proposto obteve *F1-score* de 0,9846 no dataset COVID-19 fake news e *F1-score* de 0,9101 no conjunto de dados FakeNewsNet.

*T25 - Fake News Detection on COVID-19 Tweeter Data Using Machine Learning Models* [Chakraborty et al., 2022]: O artigo propõe um modelo de detecção de notícias falsas no Twitter usando uma abordagem *ensemble*, onde as classificações geradas pelos algoritmos SVM, NB, LR e RF são combinadas a utilizando a votação majoritária. Os *datasets* utilizados são o LIAR e um dataset com tweets sobre COVID-19 disponibilizados no portal Kaggle. O pré-processamento foi realizado utilizando o NLTK para limpeza de texto em seguida foram extraídas features estatística utilizando as técnicas Chi2 e selectBest e a vetorização do conteúdo textual foi feita utilizando a técnica glove. O modelo proposto obteve melhor desempenho comparado aos algoritmos AM isolados, com acurácia de 90% no dataset COVID-19 Tweets e 70% no dataset LIAR.

*T26 - Interaction Based Credibility Analysis of News on Facebook Using Machine Learning Methodologies* [Sharmin et al., 2022]: O artigo propõe uma abordagem de experimentação utilizando algoritmos clássicos e ensembles para análise de notícias do facebook, utilizando os algoritmos: KNN, DT, SVM, LR, RF, Bagging com DT, Gradient Boost, XGBoost e Stacking utilizando DT, RF, XGboost e KNN. Os *datasets* utilizados para treinamento e teste do modelo foram coletados do Facebook, contendo notícias no idioma Bengali, divididas em três categorias: verdadeiras, falsas e satíricas. A técnica de extração de *features* utilizada foi a criação de uma pontuação das reações baseada em uma curva com dados de interação ao longo do tempo, além de outras *features* numéricas agregadas a partir dos comentários e dados das postagens. No geral os modelos *ensemble* obtiveram melhor desempenho quando comparados com os modelos clássicos de AM. O melhor desempenho foi obtido pelo XGBoost com acurácia de 88%.

*T27 - An Ensemble Learning Approach for COVID-19 Fact Verification* [Liu and Zhou, 2022]: O artigo propõe uma abordagem *ensemble* para para verificação de notícias relacionadas ao COVID-19, para isso foram criados os modelos denominados SAE, FWE e DWE, nos quais a classificação final é feita a partir da agregação das saídas dos modelos base utilizando a média simples, média ponderada e média dinamicamente ponderada respectivamente. O *dataset* utilizado é composto por três diferentes *datasets*: CoAID, FakeCovid e Zenodo, totalizando 14560 notícias e artigos em inglês

relacionados ao COVID-19. As técnicas de extração de *features* utilizadas foram CountVectorizer e TF-IDF. Os modelos ensembles gerados foram comparados vários modelos, incluindo Gradient Boost, DT, RF, DeBERTa e XLNet. O modelo proposto denominado DWE, que utiliza média dinamicamente ponderada, obteve o melhor desempenho, com uma acurácia de 98%.

*T28 - Towards a soft three-level voting model (Soft T-LVM) for fake news detection* [Jlifi et al., 2022]: O artigo apresenta o modelo Soft Three-Level Voting Model (Soft T-LVM) para detecção automática de notícias falsas relacionadas à COVID-19, onde somente a primeira camada de votação obteve os dados com as combinações de *features* extraídas dos dados e as demais foram enriquecidas com as classificações obtidas em cada fase de votação *soft*. O dataset utilizado foi o Covid-19 fake news contendo 10.700 notícias em inglês. O pré-processamento dos dados foi realizado através das técnicas de tokenização, *stemming*, lematização, remoção de pontuação e *lower casing*. As técnicas de extração de features utilizada foram a TF-IDF, Bag-of-words e N-grams. Foram comparados diversos modelos de aprendizado de máquina, incluindo SVM, RF, LR, KNN, MNNB, XGB e *voting Classifier*. No geral, o modelo proposto denominado Soft T-LVM obteve o melhor desempenho quando comparado com os demais modelos, apresentando uma acurácia final de 94,43%.

*T29 - Analysis of Deep Ensemble Transformer Model for Fake News Detection* [Panda and Kumari, 2022]: O artigo propõe o modelo ensemble baseado em em *soft voting* denominado R-X-L que é composto pelos modelos base XLM-RoBERTa, ALBERT e XLNet. O *dataset* utilizado foi obtido no kaggle e possui 25.514 amostras com os atributos: ID único, título, autor e o texto da notícia. O pré-processamento foi realizado utilizando BERT para geração dos *embeddings*. Os modelos gerados a partir dos algoritmos LSTM, GRU, XLM-RoBERTa, XLNet e ALBERT foram comparados com o modelo ensemble proposto, que obteve o maior desempenho apresentando acurácia de 98,41%.

*T30 - Fake News Detection in Social Media using Two Layers Ensemble Model* [Wynne and Swe, 2022]: O artigo apresenta um modelo de classificação de fake news baseado em *soft voting* em duas camadas, a primeira camada é composto por dois modelos de votação *soft* contendo os algoritmos RF, SVM, SGD, LR e MMNB no primeiro modelo e os algoritmos Extre Tree, Bagging, Gradient Boosting, KNN e MLP. A classificação final é dada pela combinação da saídas dos dois modelos de gerados na primeira camada utilizando a votação *soft*. O *dataset* utilizado foi o LIAR e a extração de *features* foi realizada utilizando as técnicas de Count Vectorizer e TF-IDF. O modelo *ensemble* proposto obteve o maior desempenho, apresentando 63,1% de acurácia comparado aos demais modelos executados isoladamente.

*T31 - Detecção de Fake News baseada em sinais explícitos e implícitos de um crowd híbrido: uma abordagem inspirada em meta-aprendizagem* [Souza Freire et al., 2021]: Este trabalho apresenta um modelo de detecção de fake news denominado *Hybrid Crowd Signals* (HCS), uma abordagem que utiliza a reputação dos usuários membros de um Crowd para ponderar as opiniões (*Signals*) sobre

as notícias analisadas. Esta abordagem se diferencia das abordagens *crowd signals* na forma de obter as opiniões dos usuários, ao invés de obter as opiniões fornecidas pelos usuários de forma explícita nas redes sociais, a HCS obtém opiniões implícitas a partir de padrões de comportamento dos usuários ao publicar ou compartilhar notícias, considerando o fato de um usuário divulgar uma notícia como um sinal implícito de que o mesmo considera a notícia como verdadeira.

Além das opiniões implícitas obtidas pelos usuários, o trabalho também faz uso de predições realizadas por algoritmos de aprendizado de máquina, tratando-as como opiniões explícitas das máquinas, formando assim um *Crowd Híbrido* composto tanto pelas opiniões implícitas dos usuários quanto pelas opiniões explícitas provenientes dos modelos de aprendizado de máquina.

Nos experimentos realizados, foram utilizados cinco diferentes *datasets* e diferentes modelos de aprendizado de máquina, de acordo com o tipo de dado presente em cada *dataset*. Os *dataset* Gossip e Polifact foram obtidos do repositório FakeNewsNet2; ambos possuíam notícias em inglês com dados de propagação (i.e. contagem de likes, replies e retweets) do Twitter e nome do site fonte utilizado para rotular cada notícia. O *dataset* Gossip2 é um subconjunto do *dataset* Gossip, o *dataset* FakeNewsSet foi obtido a partir de um framework para criação de *dataset* para detecção de fake news, contendo dados de propagação obtidos no Twitter para notícias em português e o *dataset* FakeBr contém notícias em português divulgadas em diversos meios digitais, possuindo apenas os dados de publicação. Na imagem III.1, é possível observar os dados estatísticos dos *datasets* utilizados.

Dataset	Notícias Not Fake	Notícias Fake	Usuários	Notícias por usuário	Usuários por notícia	
FakeNewsNet2	Gossip	5000	5000	279059	3,69	103,02
	PolitiFact	300	300	284342	1,75	832,89
	Gossip2	1200	1200	268738	2,66	298,49
FakeNewsSet	300	300	16024	1,68	45,09	
FakeBr	3600	3600	29	248,27	1	

Figura III.1: Visão estatísticas dos datasets utilizados [Freire and Goldschmidt, 2019]

Na abordagem HCS foram implementados dois métodos: o HCS-I, que se baseou exclusivamente nas opiniões implícitas dos usuários, e o método HCS-F, que fez uso tanto das opiniões implícitas dos usuários quanto das opiniões explícitas das máquinas. As opiniões explícitas das máquinas foram obtidas por meio dos algoritmos RF, XGBoost, SVM DMtext e FNE. Cabe destacar que o cálculo das reputações foi aplicado tanto aos usuários quanto às máquinas presentes na comunidade (crowd).

Para obtenção da classificação final das notícias, um combinador agrega as opiniões dos membros do crowd com as suas respectivas reputações, calculadas tanto para usuários humanos quanto para as máquinas selecionadas. Dentre os métodos testados, o HCS-F obteve melhor desempenho comparado ao HCS-I e demais modelos de aprendizado executados isoladamente, os resultados dos métodos HCS e dos algoritmos de aprendizado utilizados podem ser contemplados na imagem III.2.

Dataset	Método	Acurácia	Precisão	Recall	F1
Gossip	SL_RF	0.9304 ± 0.0083	0.9556 ± 0.0100	0.9009 ± 0.0134	0.9274 ± 0.0095
	SL_XGBOOST	0.9308 ± 0.0094	0.9653 ± 0.0112	0.8920 ± 0.0151	0.9272 ± 0.0109
	SL_SVM	0.9083 ± 0.0094	0.9651 ± 0.0120	0.8446 ± 0.0173	0.9008 ± 0.0131
	HCS-I	0.9389 ± 0.0094	0.9937 ± 0.0031	0.8822 ± 0.0205	0.9345 ± 0.0106
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM)	<b>0.9671 ± 0.0094</b>	<b>0.9940 ± 0.0029</b>	<b>0.9391 ± 0.0207</b>	<b>0.9657 ± 0.0100</b>
PolitiFact	SL_RF	0.7134 ± 0.0802	0.7041 ± 0.0859	0.7485 ± 0.0943	0.7220 ± 0.0723
	SL_XGBOOST	0.6884 ± 0.0637	0.6675 ± 0.1062	0.7456 ± 0.1047	0.7000 ± 0.0908
	SL_SVM	0.6075 ± 0.0881	0.5694 ± 0.0838	<b>0.9344 ± 0.0890</b>	0.7027 ± 0.0713
	HCS-I	0.9013 ± 0.0484	0.9761 ± 0.0391	0.8283 ± 0.0897	0.8927 ± 0.0506
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM)	<b>0.9109 ± 0.0506</b>	<b>0.9796 ± 0.0336</b>	0.8437 ± 0.0916	<b>0.9034 ± 0.0537</b>
Gossip2	SL_RF	0.8079 ± 0.0228	0.8311 ± 0.0383	0.7814 ± 0.0346	0.8047 ± 0.0236
	SL_XGBOOST	0.8140 ± 0.0344	0.8501 ± 0.0342	0.7730 ± 0.0572	0.8082 ± 0.0319
	SL_SVM	0.6279 ± 0.0324	0.7910 ± 0.0707	0.3654 ± 0.0398	0.4983 ± 0.0402
	HCS-I	0.9078 ± 0.0238	0.9708 ± 0.0218	0.8444 ± 0.0395	0.9027 ± 0.0255
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM)	<b>0.9104 ± 0.0249</b>	<b>0.9731 ± 0.0236</b>	<b>0.8473 ± 0.0436</b>	<b>0.9052 ± 0.0288</b>
FakeNewsSet	SL_RF	0.8389 ± 0.0631	0.8379 ± 0.0909	0.8449 ± 0.0887	0.8377 ± 0.0860
	SL_XGBOOST	0.8117 ± 0.0864	0.8375 ± 0.0884	0.7699 ± 0.1287	0.7983 ± 0.0953
	SL_SVM	0.6295 ± 0.0818	0.9264 ± 0.1202	0.0288 ± 0.1006	0.4303 ± 0.1215
	DMText	0.8868 ± 0.0432	0.9300 ± 0.8541	0.9405 ± 0.0588	0.8930 ± 0.0356
	FNE	0.9098 ± 0.0423	0.9521 ± 0.8709	<b>0.9629 ± 0.0551</b>	0.9123 ± 0.0440
	HCS-I	0.9179 ± 0.0397	<b>0.9832 ± 0.0404</b>	0.8542 ± 0.0852	0.9109 ± 0.0424
	HCS-F (SL_RF,SL_XGBOOST, SL_SVM, DMText, FNE)	<b>0.9639 ± 0.0281</b>	0.9831 ± 0.0274	0.9459 ± 0.0511	<b>0.9631 ± 0.0250</b>
	DMText	0.9330 ± 0.0109	0.9354 ± 0.0139	0.9296 ± 0.0101	0.9325 ± 0.0108
FakeBr	FNE	0.9339 ± 0.0105	0.9315 ± 0.0126	0.9364 ± 0.0113	0.9339 ± 0.0093
	HCS-I	0.9987 ± 0.0028	<b>0.9986 ± 0.0045</b>	0.9990 ± 0.0032	0.9988 ± 0.0026
	HCS-F (DMText, FNE)	<b>0.9989 ± 0.0026</b>	0.9980 ± 0.0046	<b>1.0000 ± 0.0000</b>	<b>0.9990 ± 0.0023</b>

Figura III.2: Resultados das máquinas x HCS em cada *dataset* [Souza Freire et al., 2021]

Os trabalhos analisados apresentam uma variedade de abordagens para detecção de fake news em diferentes formatos. No contexto deste estudo, serão consideradas para a experimentação as abordagens baseadas na utilização de atributos textuais, de propagação ou publicação das notícias que estejam disponíveis nos datasets que serão utilizados. Foi possível observar que, dentre os trabalhos listados, nenhum método *ensemble* foi aplicado para detecção de fake news baseado em crowd signals explícitos ou implícitos. Por este motivos, é esperado que o presente trabalho alcance uma nova abordagem que tenha desempenho equiparável ou superior aos demonstrados no trabalho base. A seguir, a Tabela III.1 consolida as abordagens de comitê observadas nos trabalhos relacionados.

Trabalho - Autor	Comitê	Crowd Signals
T1 - Ayyub et al. [2022]	Votação	Não
T2 - Biradar et al. [2022]	Votação	Não
T3 - Sandrilla and Devi [2022]	Votação	Não
T4 - Meel and Vishwakarma [2021]	Votação Máxima	Não
T5 - Ganesh et al. [2021]	Votação	Não
T6 - Rahman et al. [2021]	Votação Máxima	Não
T7 - Priya and Kumar [2021]	Função Softmax	Não
T8 - Castillo et al. [2021]	Bagging	Não
T9 - Ramkissoon and Goodridge [2021]	Gradient Boosting	Não
T10 - Hansrajh et al. [2021]	Blending	Não
T11 - Khan and Alhazmi [2020]	AdaBoost	Não
T12 - Sangamnerkar et al. [2020]	Bagging	Não
T13 - Kaliyar et al. [2019]	Gradient Boosting	Não
T14 - Gravanis et al. [2019]	AdaBoost	Não
T15 - Rezaei et al. [2022]	Stacking	Não
T16 - Gethsia and Juliet [2023]	Votação	Não
T17 - Varshney and Vishwakarma [2023]	Votação	Não
T18 - Zhan et al. [2022]	XGboost	Não
T19 - Alghamdi et al. [2022]	Votação	Não
T20 - Vasist and Sebastian [2022]	Votação Máxima	Não
T21 - Al Obaid et al. [2022]	Votação Majoritária Ponderada	Não
T22 - Ali et al. [2022]	MLP	Não
T23 - Imbwaga et al. [2022]	XGboost	Não
T24 - Das et al. [2022]	Votação	Não
T25 - Chabukswar et al. [2022]	Votação Majoritária	Não
T26 - Sharmin et al. [2022]	XGBoost	Não
T27 - Liu and Zhou [2022]	Votação	Não
T28 - Jlifi et al. [2022]	Votação Soft	Não
T29 - Panda and Kumari [2022]	Votação Soft	Não
T30 - Wynne and Swe [2022]	Votação Soft	Não
T31 - Souza Freire et al. [2021]	- - -	Sim

Tabela III.1: Consolidação dos trabalhos relacionados

## Capítulo IV Metodologia dos Experimentos e Resultados

Neste capítulo serão apresentadas as etapas de cada fase de experimentação dos modelos ensembles desenvolvidos, objetivando indícios de validade da hipótese de que é possível utilizar ensembles para construir modelos que superem o desempenho apresentado pelo método HCS-F desenvolvido por Souza Freire et al. [2021].

### IV.1 Formalização dos Ensembles desenvolvidos

Para a definição formal dos *Ensembles* desenvolvidos neste trabalho, seja um conjunto  $T$  de algoritmos de classificação em que  $T \neq \emptyset$ . Seja, ainda, um *Ensemble*  $E(C, I, M)$  voltado à tarefa de classificação, no qual  $C$  ( $C \neq \emptyset$  e  $\forall c \in C$ ,  $c$  é um modelo de classificação gerado a partir de algum  $t$ , tal que  $t \in T$ ) é o conjunto de classificadores da camada de base de  $E$ ,  $I$  é a estratégia de integração das saídas de  $C$  (i.e.,  $I \in \{\text{Combinação, Votação}\}$ ), e  $M$  é o método que implementa tal estratégia.

Para os experimentos deste trabalho,  $T = \{DMText, FNE, XGBoost, Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), Naive Bayes (NB), Multilayer Perceptron (MLP), Logistic Regression (LR)\}^1$ . As escolhas por esses algoritmos remetem a opções clássicas na área de *ML*, com boa diversidade em seus vieses de busca, ou por terem um histórico bem sucedido na detecção de *Fake News* (como no caso do *DMText* [Moraes et al., 2019] e do *FNE* [de Souza et al., 2020]).

### IV.2 Fase 1 - Comparativo entre os ensembles desenvolvidos e o método HCS-I

A primeira fase, teve como objetivo gerar modelos *Ensembles* que combinam as opiniões explícitas das máquinas da abordagem HCS e gerar variações dos modelos iniciais a partir das técnicas de *bagging* (utilizando os modelos gerados inicialmente como classificadores de base) e votação (utilizando os melhores classificadores encontrados nas etapas anteriores). Os resultados obtidos foram utilizados no comparativo com o método HCS-I (método que utiliza somente as opiniões implícitas dos usuários), objetivando encontrar evidências iniciais da plausibilidade da hipótese levantada neste trabalho. As próximas subseções detalham cada etapa da primeira fase de experimentos, contempladas graficamente na Figura IV.24.

<sup>1</sup>Code Link: <https://github.com/Uriel-Merola/EHCS/tree/main/Ensembles>

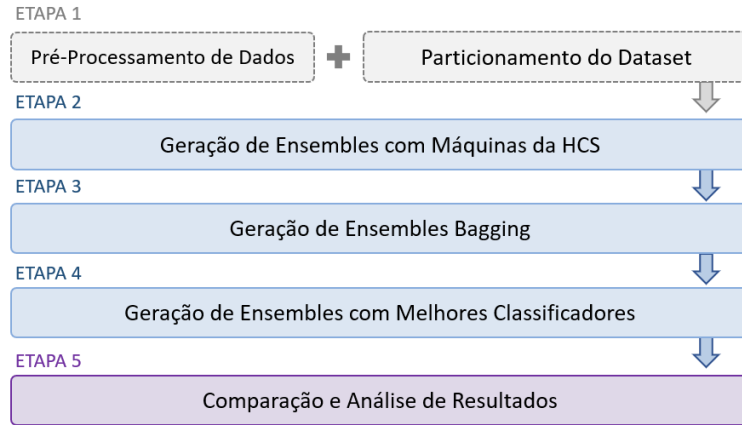


Figura IV.1: Etapas da metodologia de experimentação adotada na fase 1

De forma a comparar os *Ensembles* desenvolvidos neste trabalho com os métodos HCS-I e HCS-F do trabalho base, foram utilizados nos experimentos os mesmos *datasets*<sup>2</sup> empregados em Souza Freire et al. [2021]. A Tabela IV.1 apresenta um resumo estatístico desses conjuntos de dados. Cada um deles possui notícias com as seguintes informações: *id* (identificador da notícia), *rótulo* (sendo  $\text{rótulo} \in \{\text{real}, \text{fake}\}$ ), e *classificações das máquinas*<sup>3</sup>. As máquinas utilizadas em cada *dataset* também encontram-se indicadas na Tabela IV.1.

<i>Datasets</i>	Notícias <i>Fake</i>	Notícias Não <i>Fake</i>	Máquinas utilizadas em Souza Freire et al. [2021]
Gossip	5000	5000	<i>Random Forest</i> , <i>XGBoost</i> e <i>SVM</i>
PolitiFact	300	300	<i>Random Forest</i> , <i>XGBoost</i> e <i>SVM</i>
Gossip2	1200	1200	<i>Random Forest</i> , <i>XGBoost</i> e <i>SVM</i>
FakeNewsSet	300	300	<i>Random Forest</i> , <i>XGBoost</i> e <i>SVM</i> , <i>DMText</i> e <i>FNE</i>
FakeBr	3600	3600	<i>DMText</i> e <i>FNE</i>

Tabela IV.1: Síntese dos *Datasets* utilizados nos experimentos deste trabalho.

#### IV.2.1 Pré-Processamento e Particionamento do Dataset na Fase 1

Na Etapa 1, o pré-processamento envolveu a conversão dos valores alfanuméricos dos atributos categóricos em valores numéricos (i.e.  $\text{fake} \rightarrow 0$  e  $\text{real} \rightarrow 1$ ). Tal conversão visou garantir a compatibilidade dos valores dos atributos com as demandas de algoritmos de  $T$  em relação aos tipos de dados. Passaram por essa conversão os valores dos atributos *rótulo* e *classificações das máquinas*.

Ainda na Etapa 1, cada *dataset* foi particionado em 50% dos dados para teste e 50% dos dados para treinamento dos modelos de classificação. A fim de viabilizar uma comparação justa dos *Ensembles* deste trabalho com o método HCS-I, a divisão de cada *dataset* foi a mesma realizada em Souza Freire et al. [2021].

<sup>2</sup>Dataset Link: <https://github.com/Uriel-Merola/EHCS/tree/main/Datasets>

<sup>3</sup>No contexto do método HCS-F, uma máquina é um classificador de notícias como *real* ou *fake*.

### IV.2.2 Geração de Ensembles com Máquinas do HCS-F

A Etapa 2 teve como objetivo construir *Ensembles*  $E_j(C, I, M)$ , nos quais  $C$  fosse formado por modelos de classificação gerados a partir das mesmas máquinas utilizadas pelo método HCS-F (i.e.,  $\forall c \in C, \exists t \in T'$ , tal que  $c$  foi gerado por  $t$ , sendo  $T' = \{DMText, FNE, SVM, RF, XGBoost\}$ ) e  $I=Combinação$ . As especificações dos *Ensembles* gerados estão indicadas na Tabela IV.2.

<i>Ensembles</i>	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$
Especificações	$M = KNN$	$M = DT$	$M = NB$	$M = SVM$	$M = MLP$	$M = LR$	$M = RF$

Tabela IV.2: Especificações dos *Ensembles* construídos com máquinas do HCS-F

Uma vez definidos os *Ensembles* a serem implementados, foi executado o processo de validação cruzada com 10 conjuntos de cada *Ensemble* aplicado ao conjunto de testes. Para tanto, foram adotados os valores *default* dos hiperparâmetros dos algoritmos de  $C$ . O desempenho dos *Ensembles* com máquinas do HCS-F executados em cada dataset pode ser contemplado nos gráficos abaixo, bem como a consolidação dos resultados obtidos (acurácia e desvio padrão) apresentada na tabela IV.3.

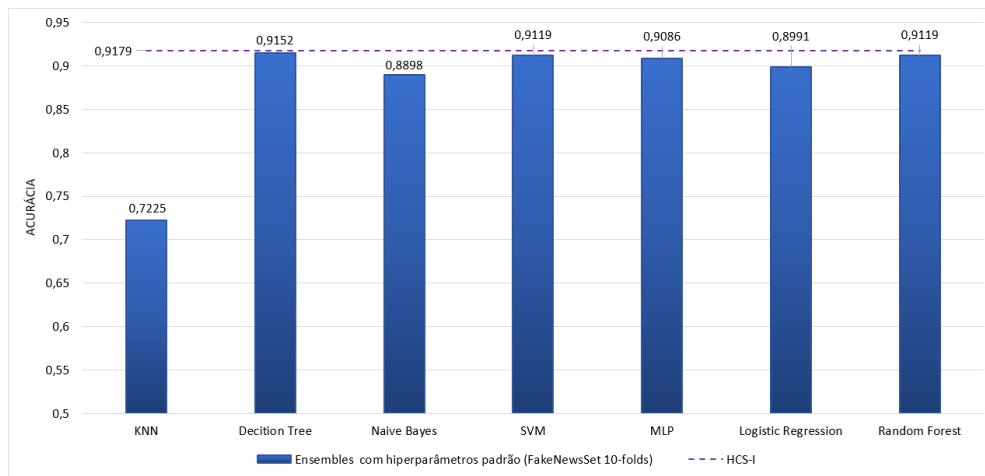


Figura IV.2: Acurácia dos *Ensembles* com Máquinas do HCS-F no dataset FakeNewsSet



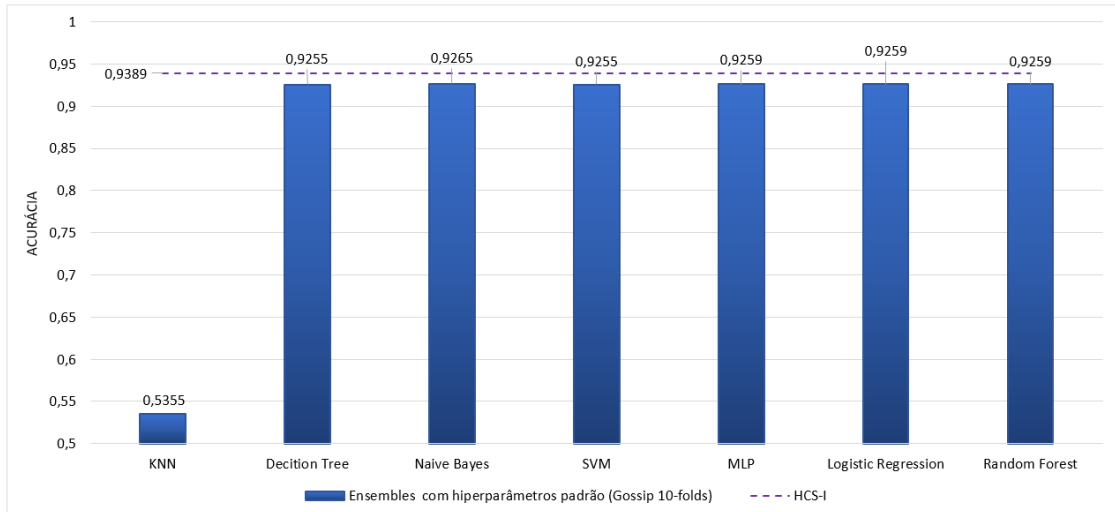


Figura IV.3: Acurácia dos *Ensembles* com Máquinas do HCS-F no *dataset* Gossip

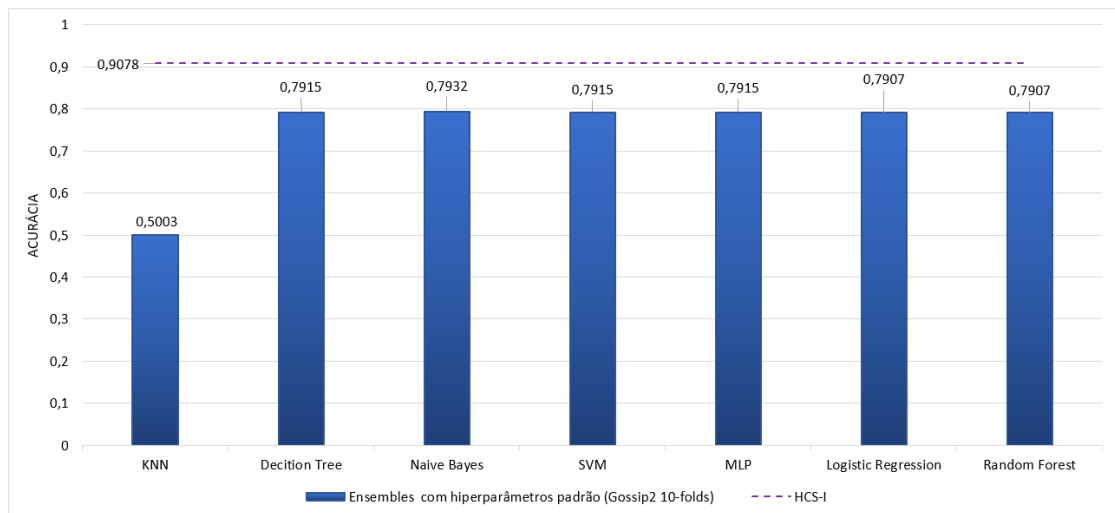


Figura IV.4: Acurácia dos *Ensembles* com Máquinas do HCS-F no *dataset* Gossip2

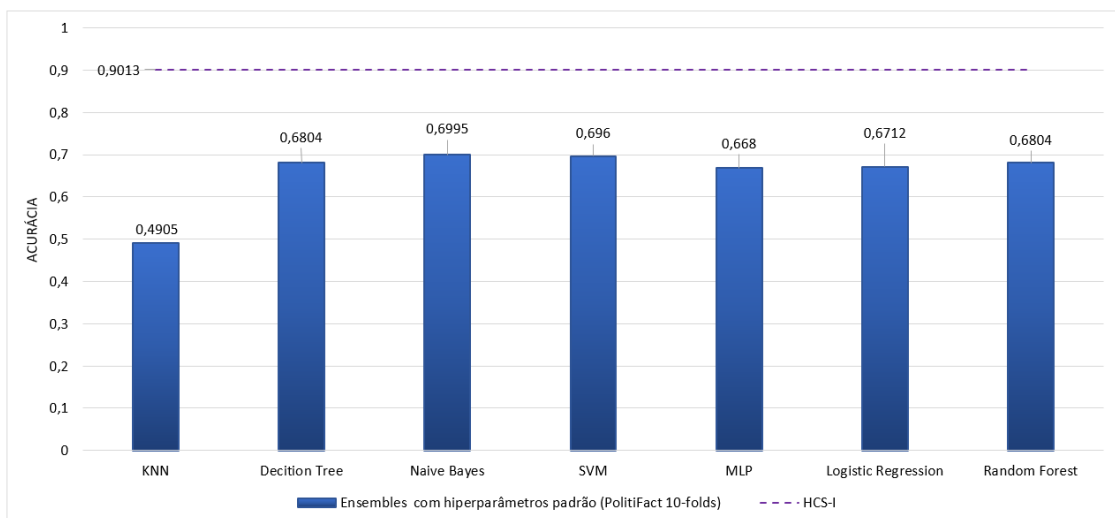


Figura IV.5: Acurácia dos *Ensembles* com Máquinas do HCS-F no *dataset* PolitiFact

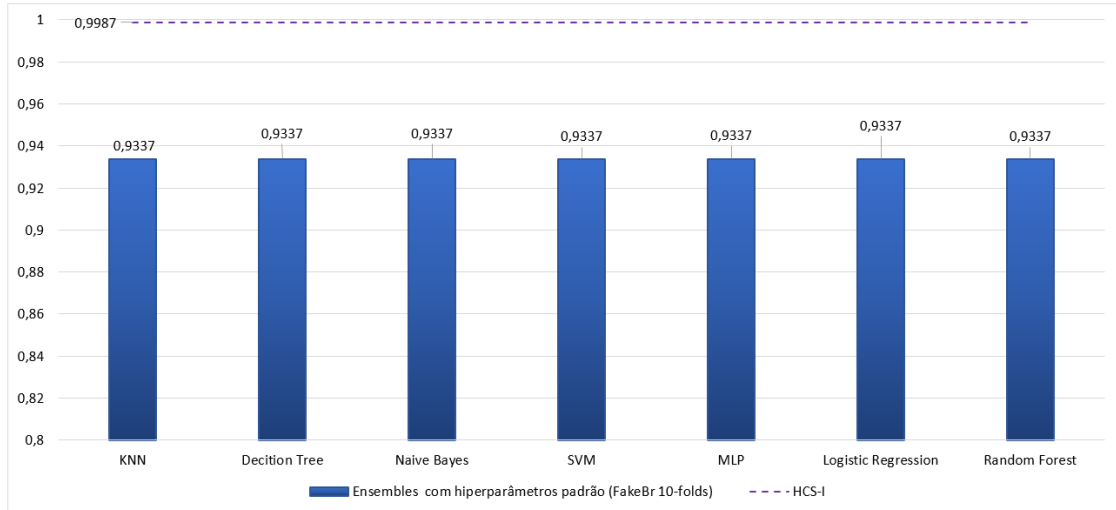


Figura IV.6: Acurácia dos *Ensembles* com Máquinas do HCS-F no *dataset* FakeBr

Modelos	FakeNewsSet	Gossip	Gossip2	PolitiFact	FakeBr
HCS-I	<b>0.9179± 0.0397</b>	<b>0,9389± 0.0094</b>	<b>0,9078±0.0238</b>	<b>0,9013± 0.0484</b>	<b>0,9987± 0.0028</b>
$E_1$	0.7225±0.0600	0.5355±0.0900	0.5003± 0.0294	0.4905±0.0144	0.9337± 0.0095
$E_2$	0.9152±0.0628	0.9255±0.1423	0.7915±0.1799	0.6804±0.1078	0.9337± 0.0095
$E_3$	0.8898±0.0635	0.9265±0.1420	0.7932±0.1826	0.6995±0.1227	0.9337± 0.0095
$E_4$	0.9119±0.0611	0.9255±0.1423	0.7915±0.1799	0.6960±0.0957	0.9337± 0.0095
$E_5$	0.9086±0.0576	0.9259±0.1424	0.7915±0.1799	0.6680±0.1199	0.9337± 0.0095
$E_6$	0.8991±0.0563	0.9259±0.1424	0.7907±0.1752	0.6712±0.1254	0.9337± 0.0095
$E_7$	0.9119±0.05913	0.9259±0.1425	0.7907±0.1811	0.6804±0.1078	0.9337± 0.0095

Tabela IV.3: Acurácia média  $\pm$  desvio-padrão dos *Ensembles* com máquinas do HCS-F no processo de validação cruzada com 10 conjuntos em cada *dataset* (Fase 1)

### IV.2.3 Geração de Ensembles Bagging da Fase 1

A fim de variar os tipos de *Ensemble* criados na Etapa 2, a Etapa 3 consistiu em gerar *Ensembles* do tipo *bagging* a partir dos algoritmos do conjunto  $T'' \subset T$ , sendo  $T'' = \{KNN, DT, SVM, MLP, NB, LR, RF\}^4$ . Nesta etapa, foram adotados os valores *default* dos hiperparâmetros dos algoritmos de classificação em  $T''$ . É importante ressaltar que o algoritmo *bagging* também foi empregado com seus hiperparâmetros *default*, e que por definição  $I=Votação$  e  $M$  é o método que implementa a eleição por votação majoritária simples. As especificações dos *Ensembles* do tipo *bagging* gerados estão indicadas na Tabela IV.4.

<sup>4</sup>Tal escolha visou o uso de algoritmos de classificação tradicionais da área de *ML*.

<i>Ensembles</i>	Especificações
$E_8$	$C = \{c_{KNN,1}, c_{KNN,2}, \dots, c_{KNN,10}\}$
$E_9$	$C = \{c_{DT,1}, c_{DT,2}, \dots, c_{DT,10}\}$
$E_{10}$	$C = \{c_{NB,1}, c_{NB,2}, \dots, c_{NB,10}\}$
$E_{11}$	$C = \{c_{SVM,1}, c_{SVM,2}, \dots, c_{SVM,10}\}$
$E_{12}$	$C = \{c_{MLP,1}, c_{MLP,2}, \dots, c_{MLP,10}\}$
$E_{13}$	$C = \{c_{LR,1}, c_{LR,2}, \dots, c_{LR,10}\}$
$E_{14}$	$C = \{c_{RF,1}, c_{RF,2}, \dots, c_{RF,10}\}$

Tabela IV.4: Especificações dos *Ensembles bagging* implementados na fase 1

De forma análoga à Etapa 2, foi executado o processo de validação cruzada com 10 conjuntos para cada Ensemble da Etapa 3 aplicado ao conjunto de testes. O desempenho dos *Ensembles bagging* com máquinas do HCS-F executados em cada dataset pode ser contemplado nos gráficos abaixo. A consolidação dos resultados obtidos (Acurácia Média e Desvio Padrão) é apresentada na tabela IV.5.

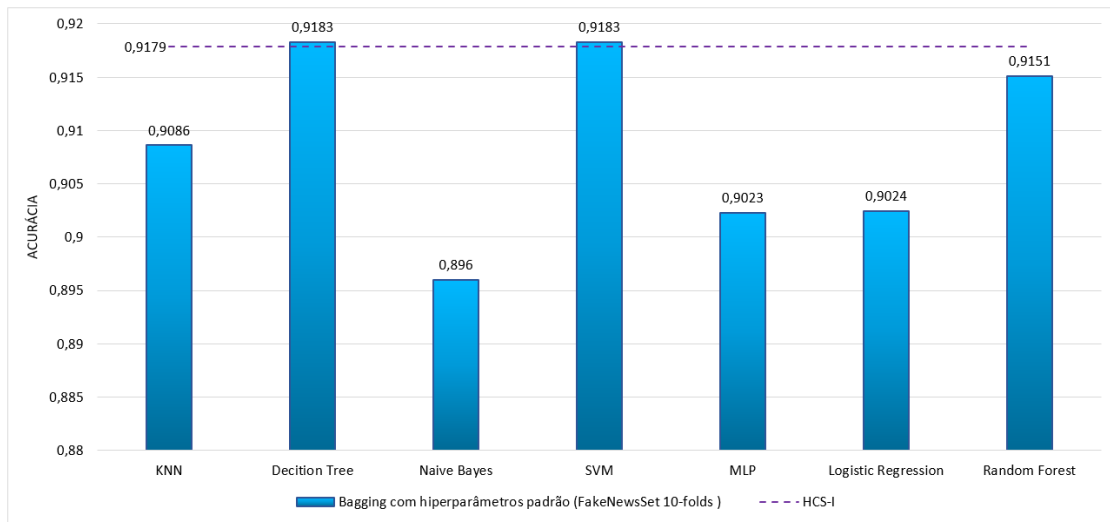


Figura IV.7: Acurácia dos *Ensembles Bagging* com Máquinas do HCS-F no *dataset* FakeNewsSet

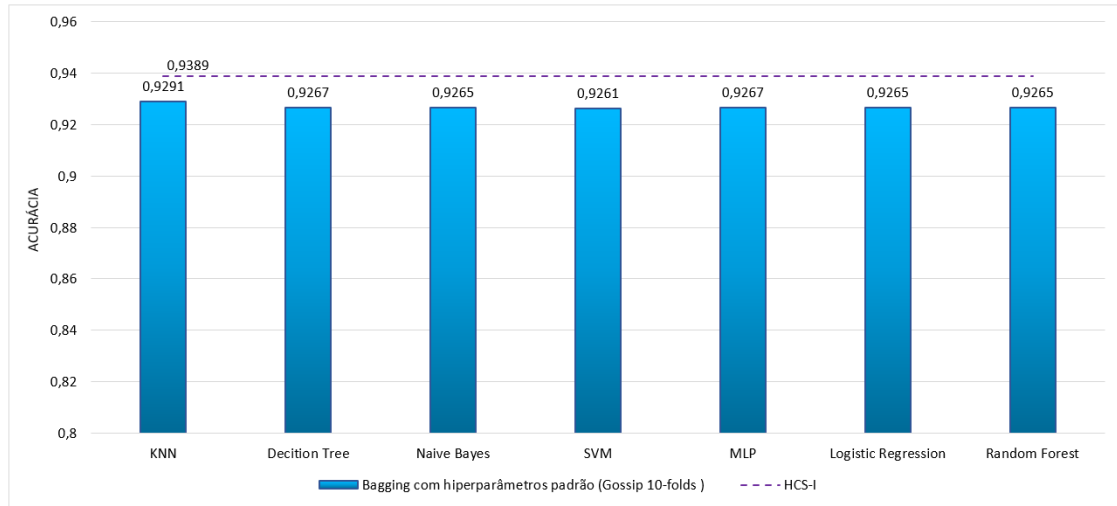


Figura IV.8: Acúrcia dos *Ensembles Bagging* com Máquinas do HCS-F no *dataset* Gossip

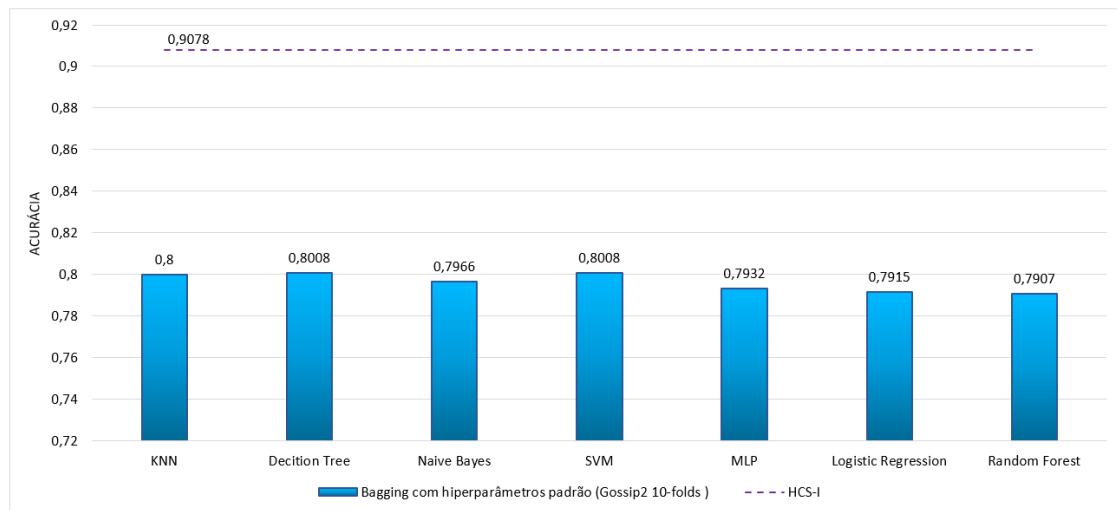


Figura IV.9: Acúrcia dos *Ensembles Bagging* com Máquinas do HCS-F no *dataset* Gossip2

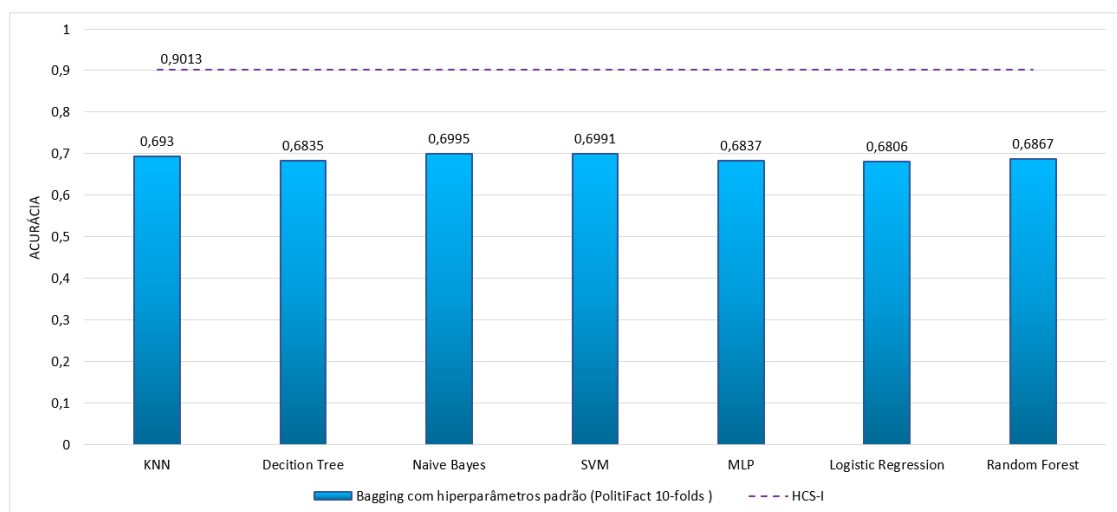


Figura IV.10: Acúrcia dos *Ensembles Bagging* com Máquinas do HCS-F no *dataset* PolitiFact

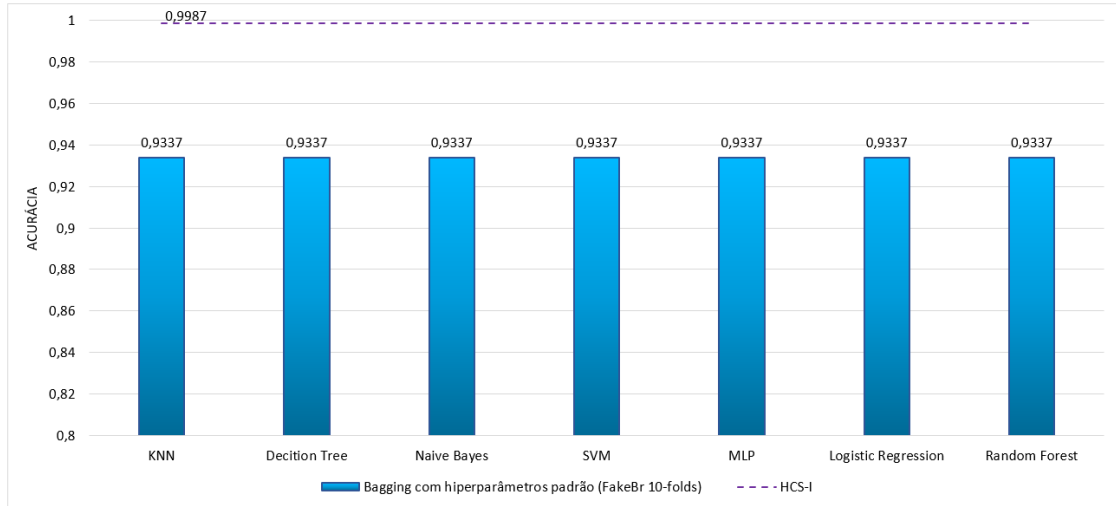


Figura IV.11: Acurácia dos *Ensembles Bagging* com Máquinas do HCS-F no *dataset* FakeBr

Modelos	FakeNewsSet	Gossip	Gossip2	PolitiFact	FakeBr
HCS-I	0.9179± 0.0397	<b>0,9389± 0.0094</b>	<b>0,9078±0.0238</b>	<b>0,9013± 0.0484</b>	<b>0,9987± 0.0028</b>
$E_8$	0.9086±0.0609	0.9291±0.1429	0.8000±0.1801	0.6930±0.0992	0.9337± 0.0095
$E_9$	<b>0.9183±0.0565</b>	0.9267±0.1429	0.8008±0.1789	0.6835±0.1111	0.9337± 0.0095
$E_{10}$	0.8960±0.0617	0.9265±0.1420	0.7966±0.1856	0.6995±0.1227	0.9337± 0.0095
$E_{11}$	<b>0.9183±0.0565</b>	0.9261±0.1426	0.8008±0.1789	0.6991±0.0989	0.9337± 0.0095
$E_{12}$	0.9023±0.0558	0.9267±0.1428	0.7932±0.1791	0.6837±0.1168	0.9337± 0.0095
$E_{13}$	0.9024±0.0592	0.9265±0.1427	0.7915±0.1762	0.6806±0.1171	0.9337± 0.0095
$E_{14}$	0.9151±0.0600	0.9265±0.1428	0.7907±0.1811	0.6867±0.1094	0.9337± 0.0095

Tabela IV.5: Acurácia média  $\pm$  desvio-padrão dos *Ensembles* bagging com máquinas do HCS-F no processo de validação cruzada com 10 conjuntos em cada *dataset*

#### IV.2.4 Geração de Ensembles com Melhores Classificadores da Fase 1

Na Etapa 4, os melhores modelos de classificação gerados para cada *dataset* nas etapas anteriores foram utilizados na formação de novos *Ensembles*, em que  $I=Votação$  e  $M$  o método de votação majoritária simples. A intenção desta escolha foi procurar tirar proveito do conhecimento sobre os *datasets* adquirido pelos referidos modelos. O desempenho obtido pelos *Ensembles* com melhores classificadores pode ser observado no gráfico abaixo. As especificações e a consolidação dos resultados obtidos em cada *dataset* estão indicadas na Tabela IV.6.

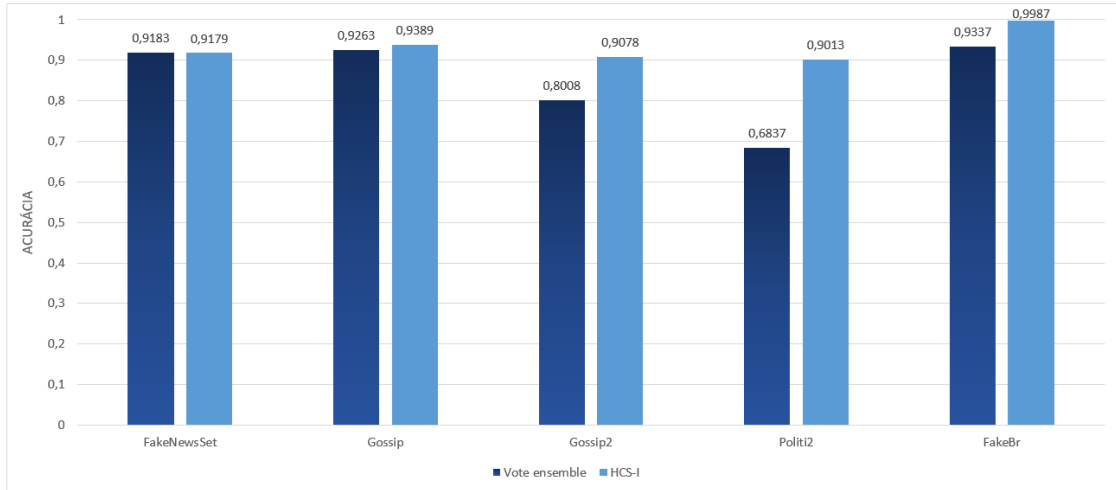


Figura IV.12: Acurácia dos *Ensembles* com Melhores Classificadores da Fase 1 em Cada *Dataset*

<i>Ensembles</i>	Especificações	Dataset	Acurácia/Desvio Padrão	HCS-I
$E_{15}$	$C = \{E_2, E_9, E_{11}\}$	FakeNewsSet	<b>0.9183</b> ± 0.0566	0.9179± 0.0397
$E_{16}$	$C = \{E_8, E_9, E_{13}\}$	Gossip	0.9263± 0.1427	<b>0,9389</b> ± <b>0.0094</b>
$E_{17}$	$C = \{E_8, E_9, E_{11}\}$	Gossip2	0.8008± 0.1790	<b>0,9078</b> ± <b>0.0238</b>
$E_{18}$	$C = \{E_9, E_{11}, E_{13}\}$	PolitiFact	0.6837± 0.1062	<b>0,9013</b> ± <b>0.0484</b>
$E_{19}$	$C = \{E_1, E_3, E_4\}$	FakeBr	0.9337± 0.0096	<b>0,9987</b> ± <b>0.0028</b>

Tabela IV.6: Especificações e resultados dos *Ensembles* com melhores classificadores - validação cruzada com 10 conjuntos em cada *dataset* (Fase 1)

#### IV.2.5 Comparação e Análise de Resultados da Fase 1

Na Etapa 5, foi realizada a comparação dos resultados gerados nos experimentos. Em geral, os *Ensembles* criados com máquinas do HCS-F (Tabela IV.2) apresentaram desempenhos inferiores aos do método HCS-I em todos os *datasets* (Tabela IV.3), indicando sinais de robustez do método que congrega opiniões dos usuários dos MDDN.

Os modelos *Ensemble* do tipo *bagging* (Tabela IV.4), em sua maioria, apresentaram um desempenho (Tabela IV.5) superior em relação aos modelos *Ensemble* construídos com as máquinas utilizadas pelo HCS-F. Tais resultados sinalizam para uma superioridade da técnica *bagging* na formação de bons *Ensembles* para detecção de *Fake News*. Por outro lado, quando comparados ao método HCS-I, os *Ensembles bagging* também apresentaram resultados inferiores, com exceção dos *Ensembles*  $E_9$  (baseado no *DT*) e  $E_{11}$  (baseado no *SVM*), que obtiveram acurácia média levemente superior no *dataset FakeNewsSet*.

De forma análoga ao comentado acima, os *Ensembles* formados a partir dos melhores classificadores obtidos nos experimentos anteriores também só conseguiram superar o método HCS-I no *dataset FakeNewsSet* (Tabela IV.6). Entretanto, excetuando-se os resultados do *Ensemble*  $E_{18}$  no *dataset PolitiFact*, os demais *Ensembles* com melhores classificadores apresentaram resultados re-

lativamente próximos aos do HCS-I, indicando que o uso de *Ensembles* pode ser uma alternativa interessante para conjugar os resultados das máquinas do HCS-F.

### IV.3 Fase 2 - Comparativo entre os ensembles desenvolvidos e método HCS-F

A segunda fase, teve como objetivo substituir a inferência bayesiana ingênua presente no método HCS-F, utilizando *Ensembles* gerados para combinar as classificações fornecidas pelas opiniões implícitas do usuários disponibilizadas pelo método HCS-I e as opiniões explícitas fornecidas pela máquinas do método HCS-F. Nessa fase, também foi realizada a variação dos *Ensembles* utilizando as técnicas de *bagging* e votação, objetivando obter melhor desempenho do que a conjugação ingênua realizada pelo método HCS-F e assim, obtendo indícios da validade da hipótese levantada. As próximas subseções detalham cada etapa da segunda fase de experimentos, contempladas graficamente na Figura IV.13.

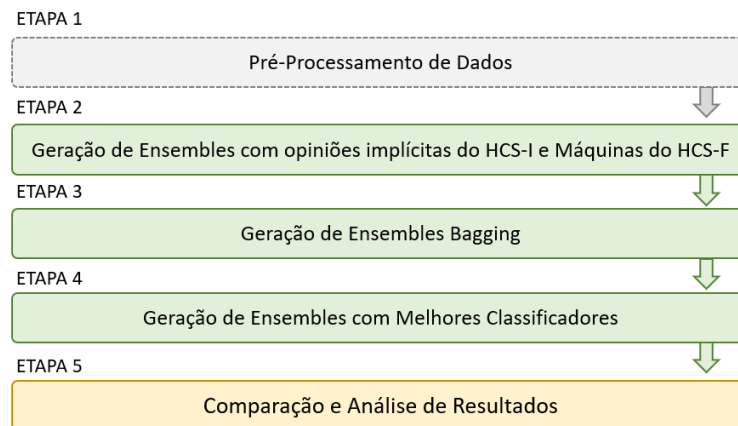


Figura IV.13: Etapas da metodologia de experimentação adotada na fase 2

#### IV.3.1 Pré-Processamento do *Dataset* na Fase 2

Na Etapa 1 da segunda fase, o pré-processamento envolveu a inclusão das classificações geradas pelas opiniões implícitas dos usuários no modelo HCS-I nos *datasets* já pré-processados e utilizados na primeira fase. As classificações do método HCS-I incluídas no dataset como atributos categóricos foram convertidas em valores numéricos (i.e. *fake*  $\rightarrow 0$  e *real*  $\rightarrow 1$ ). Tal conversão também visou garantir a compatibilidade dos valores dos atributos com as demandas de algoritmos de  $T$  em relação aos tipos de dados. As proporções de divisão do dataset foram mantidas em 50% para treinamento e 50% para testes.

### IV.3.2 Geração de Ensembles com opiniões implícitas do HCS-I e Máquinas do HCS-F

A etapa 2 da segunda fase teve como objetivo gerar *Ensembles*  $E_F(C, I, M)$ , onde  $C$  fosse formada por modelos de classificação gerados a partir das máquinas utilizadas pelo método HCS-F e pelo modelo gerado através do método HCS-I (i.e.,  $\forall c \in C, \exists t \in T'$ , tal que  $c$  foi gerado por  $t$ , sendo  $T' = \{DMText, FNE, SVM, RF, XGBoost, HCS-I\}$ ) e  $I=Combinação$ . As especificações dos *Ensembles* gerados estão indicadas na Tabela IV.7.

<i>Ensembles</i>	$E_{20}$	$E_{21}$	$E_{22}$	$E_{23}$	$E_{24}$	$E_{25}$	$E_{26}$
Especificações	$M = KNN$	$M = DT$	$M = NB$	$M = SVM$	$M = MLP$	$M = LR$	$M = RF$

Tabela IV.7: Especificações dos *Ensembles* construídos com modelos gerados por máquinas do HCS-F e opiniões implícitas do método HCS-I

Na sequência, os ensembles construídos foram executados com processo de validação cruzada em 10 conjuntos, com todos algoritmos de combinação contidos em  $C$ , utilizando hiperparâmetros *default*. Os resultados obtidos nesta etapa podem ser contemplados nos gráficos abaixo e a consolidação dos resultados obtidos em cada dataset pode ser observada na tabela IV.8.

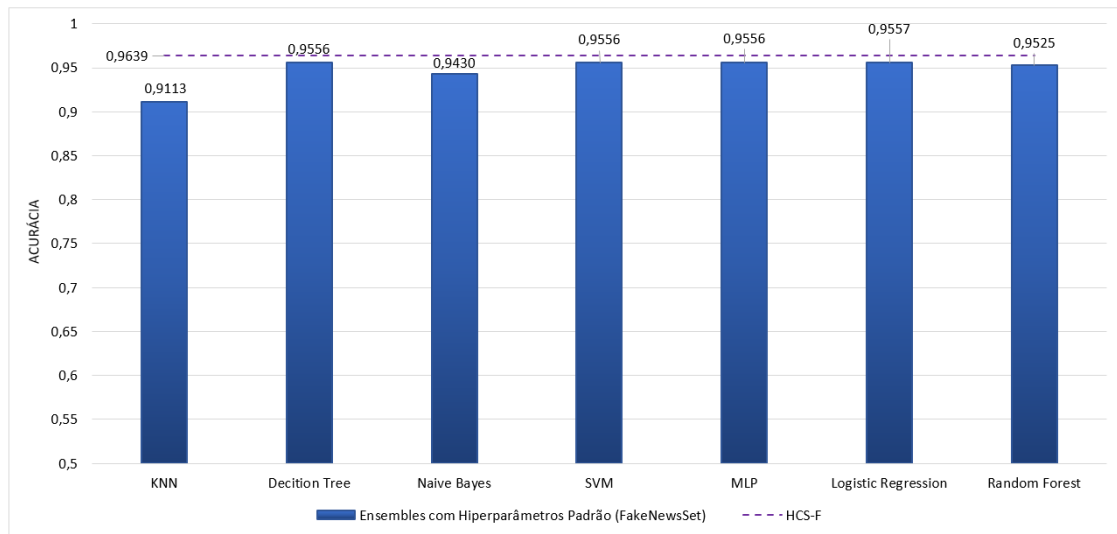


Figura IV.14: Acurácia dos *Ensembles* com opiniões implícitas do HCS-I e Máquinas do HCS-F no Dataset FakeNewsSet



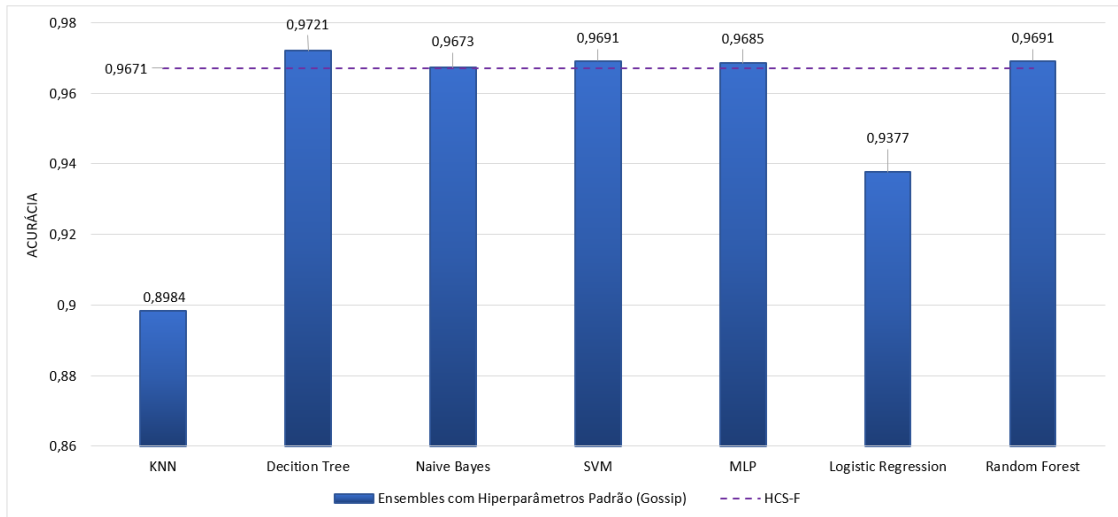


Figura IV.15: Acurácia dos *Ensembles* com opiniões implícitas do HCS-I e Máquinas do HCS-F no *Dataset Gossip*

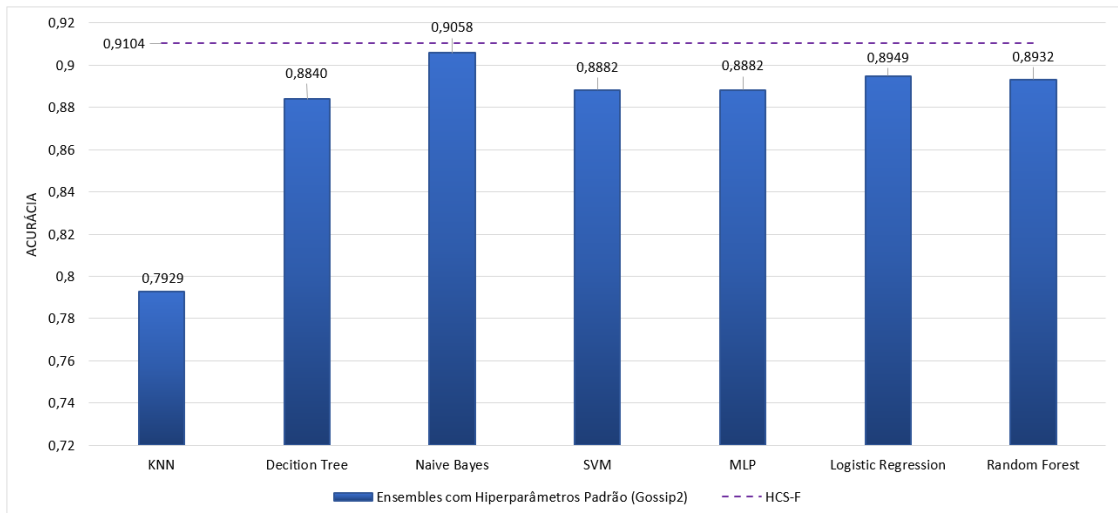


Figura IV.16: Acurácia dos *Ensembles* com opiniões implícitas do HCS-I e Máquinas do HCS-F no *Dataset Gossip2*

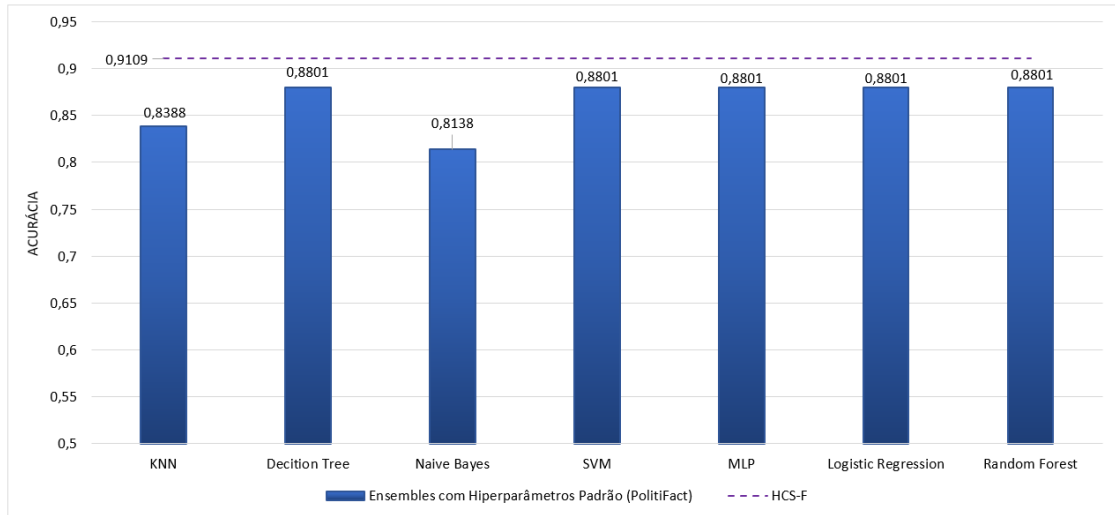


Figura IV.17: Acurácia dos *Ensembles* com opiniões implícitas do HCS-I e Máquinas do HCS-F no *Dataset* PolitiFact

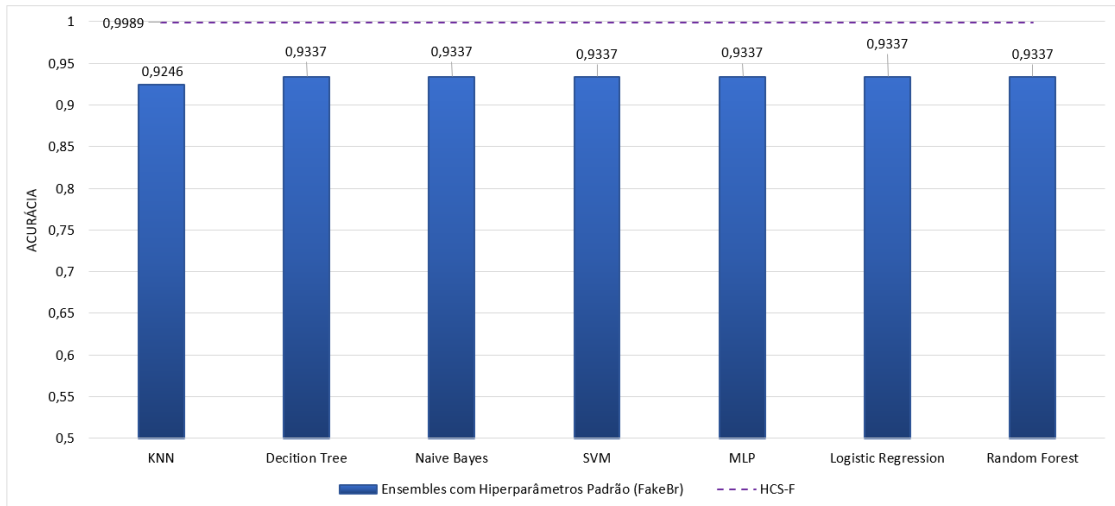


Figura IV.18: Acurácia dos *Ensembles* com opiniões implícitas do HCS-I e Máquinas do HCS-F no *Dataset* FakeBr

Modelos	FakeNewsSet	Gossip	Gossip2	PolitiFact	FakeBr
HCS-F	<b>0.9639 ± 0.0281</b>	0.9671 ± 0.0094	<b>0.9104 ± 0.0249</b>	<b>0.9109 ± 0.0506</b>	<b>0.9989 ± 0.0026</b>
$E_{20}$	0.9113 ± 0.0625	0.8984 ± 0.1130	0.7929 ± 0.1315	0.8388 ± 0.1054	0.9246 ± 0.0258
$E_{21}$	0.9556 ± 0.0454	<b>0.9721 ± 0.0462</b>	0.8840 ± 0.1363	0.8801 ± 0.0882	0.9337 ± 0.0095
$E_{22}$	0.9430 ± 0.0547	<b>0.9673 ± 0.0666</b>	0.9058 ± 0.1362	0.8138 ± 0.0927	0.9337 ± 0.0095
$E_{23}$	0.9556 ± 0.0454	<b>0.9691 ± 0.0537</b>	0.8882 ± 0.1360	0.8801 ± 0.0882	0.9337 ± 0.0095
$E_{24}$	0.9556 ± 0.0454	<b>0.9685 ± 0.0553</b>	0.8882 ± 0.1360	0.8801 ± 0.0882	0.9337 ± 0.0095
$E_{25}$	0.9557 ± 0.0431	0.9377 ± 0.1242	0.8949 ± 0.1305	0.8801 ± 0.0882	0.9337 ± 0.0095
$E_{26}$	0.9525 ± 0.0433	<b>0.9691 ± 0.0537</b>	0.8932 ± 0.1303	0.8801 ± 0.0882	0.9337 ± 0.0095

Tabela IV.8: Acurácia média  $\pm$  desvio-padrão dos *Ensembles bagging* no processo de validação cruzada com 10 conjuntos em cada *dataset* (Fase 2)

### IV.3.3 Geração de Ensembles Bagging da Fase 2

Na etapa 3 da segunda fase, foram gerados *Ensembles* do tipo *bagging* com opiniões implícitas do HCS-I e máquinas do HCS-F, usando algoritmos de base com hiperparâmetros padrão, contidos no conjunto  $T''$  sendo  $T'' = \{KNN, DT, SVM, MLP, NB, LR, RF\}$ , o método de combinação I = Votação onde é implementado a eleição por votação majoritária simples. A composição desses *Ensembles* do tipo *bagging* pode ser observada na tabela IV.9.

<i>Ensembles</i>	Especificações
$E_{27}$	$C = \{c_{KNN,1}, c_{KNN,2}, \dots, c_{KNN,10}\}$
$E_{28}$	$C = \{c_{DT,1}, c_{DT,2}, \dots, c_{DT,10}\}$
$E_{29}$	$C = \{c_{NB,1}, c_{NB,2}, \dots, c_{NB,10}\}$
$E_{30}$	$C = \{c_{SVM,1}, c_{SVM,2}, \dots, c_{SVM,10}\}$
$E_{31}$	$C = \{c_{MLP,1}, c_{MLP,2}, \dots, c_{MLP,10}\}$
$E_{32}$	$C = \{c_{LR,1}, c_{LR,2}, \dots, c_{LR,10}\}$
$E_{33}$	$C = \{c_{RF,1}, c_{RF,2}, \dots, c_{RF,10}\}$

Tabela IV.9: Especificações dos *Ensembles bagging* com opiniões implícitas do HCS-I e máquinas do HCS-F

Os *Ensembles* do tipo *bagging* gerados na segunda fase também foram executados com validação cruzada com 10 conjuntos aplicados no dataset de teste. O desempenho dos *Ensembles bagging* pode ser contemplado nos gráficos abaixo e a consolidação dos resultados (Acurácia Média e Desvio Padrão) é apresentada na tabela IV.10

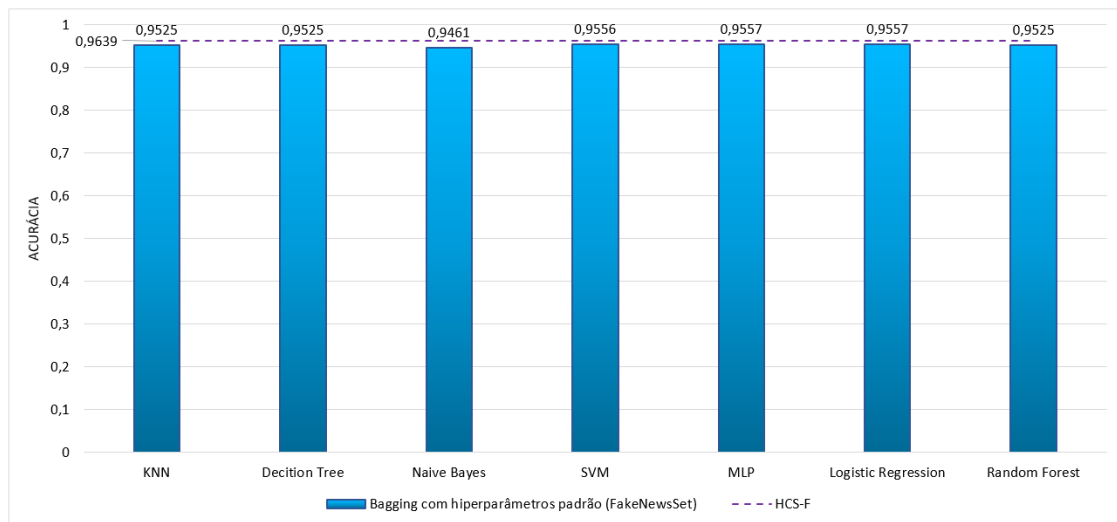


Figura IV.19: Acurácia dos *Ensembles Bagging* com opiniões implícitas do HCS-I e Máquinas do HCS-F no *Dataset FakeNewsSet*

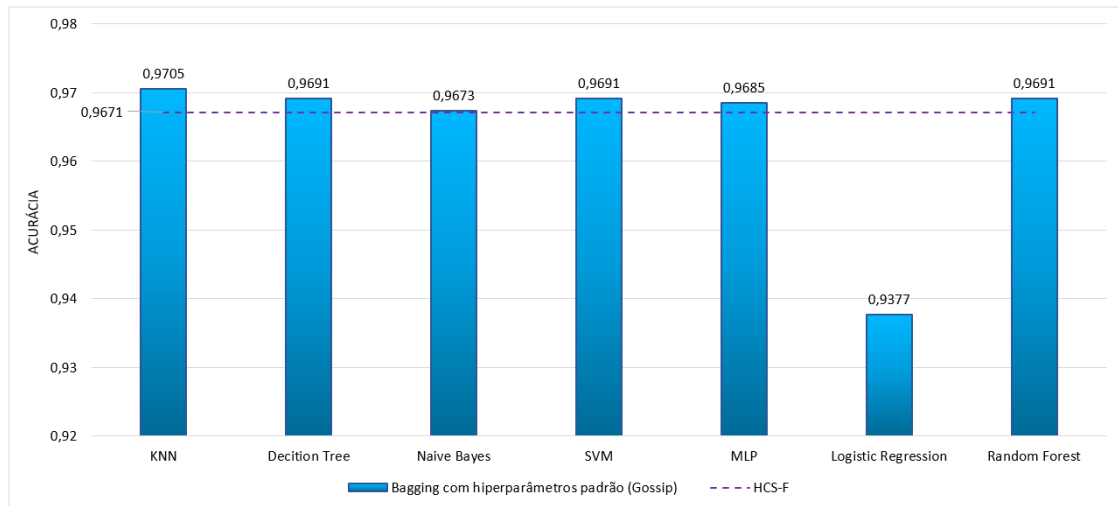


Figura IV.20: Acurácia dos *Ensembles Bagging* com opiniões implícitas do HCS-I e Máquinas do HCS-F no *Dataset Gossip*

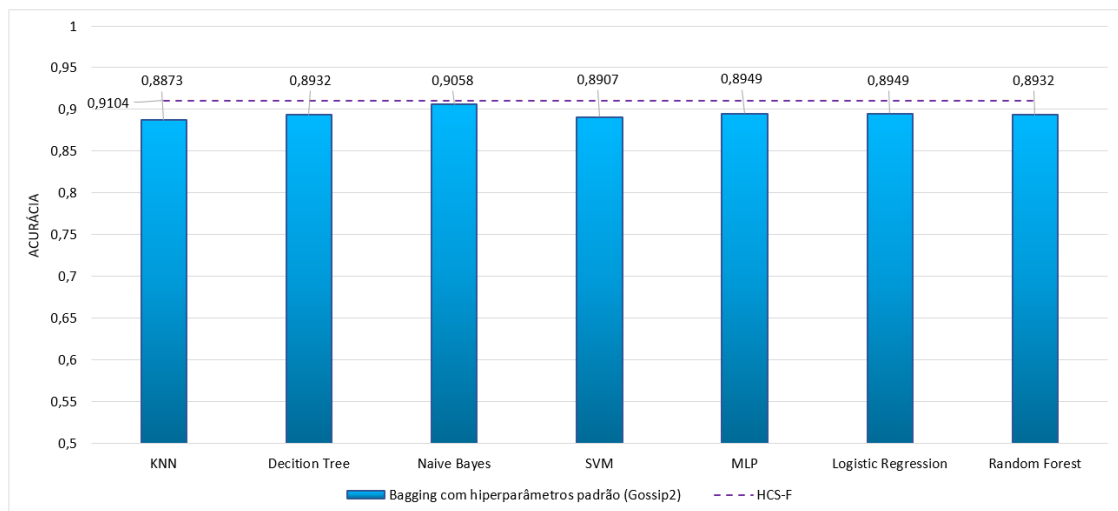


Figura IV.21: Acurácia dos *Ensembles Bagging* com opiniões implícitas do HCS-I e Máquinas do HCS-F no *Dataset Gossip2*

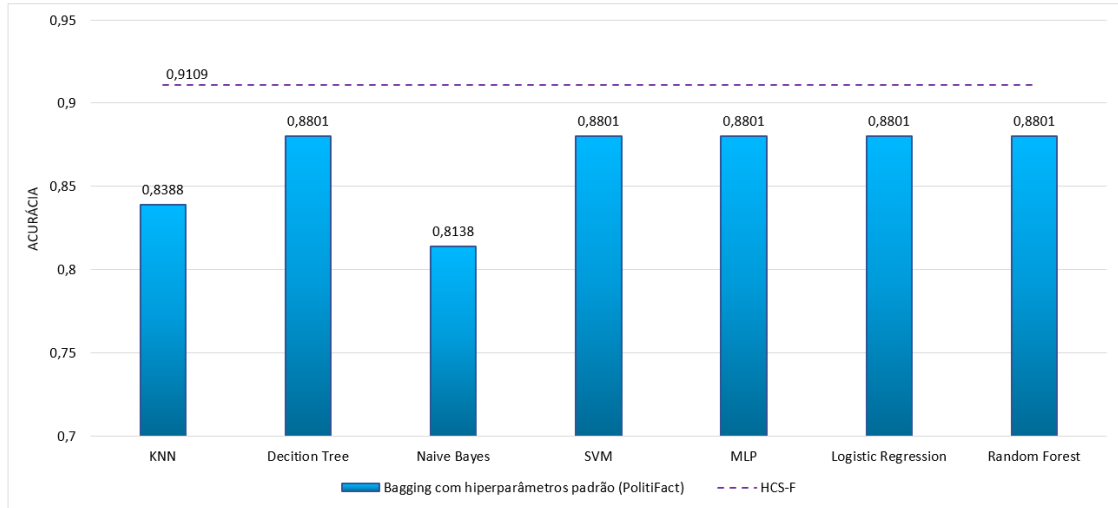


Figura IV.22: Acurácia dos *Ensembles Bagging* com opiniões implícitas do HCS-I e Máquinas do HCS-F no *Dataset PolitiFact*

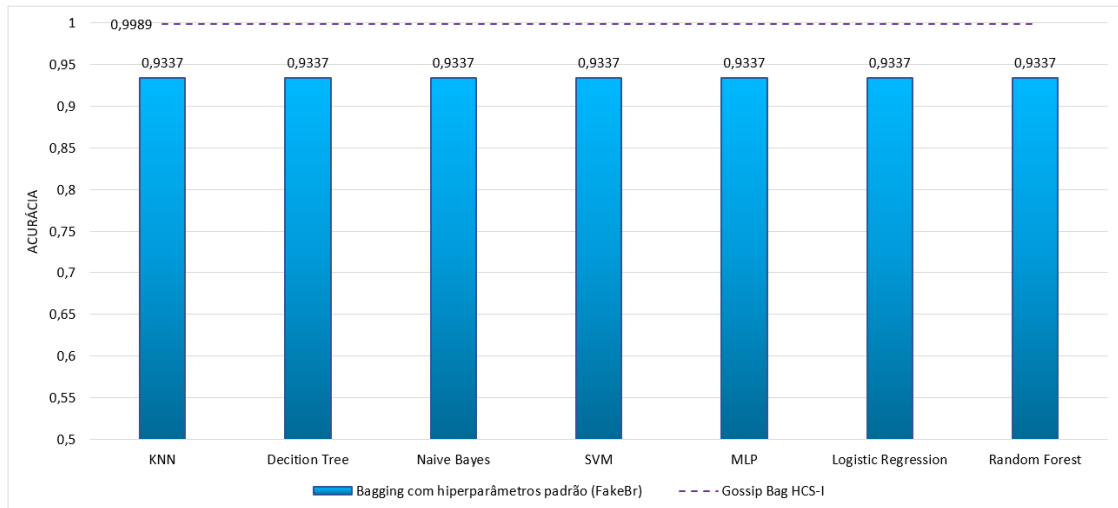


Figura IV.23: Acurácia dos *Ensembles Bagging* com opiniões implícitas do HCS-I e Máquinas do HCS-F no *Dataset FakeBr*

Modelos	FakeNewsSet	Gossip	Gossip2	PolitiFact	FakeBr
HCS-F	<b>0.9639 ± 0.0281</b>	0.9671 ± 0.0094	<b>0.9104 ± 0.0249</b>	<b>0.9109 ± 0.0506</b>	<b>0.9989 ± 0.0026</b>
$E_{27}$	0.9525 ± 0.0519	<b>0.9705 ± 0.0501</b>	0.8873 ± 0.1265	0.8801 ± 0.0882	0.9337 ± 0.0095
$E_{28}$	0.9525 ± 0.0433	<b>0.9691 ± 0.0537</b>	0.8932 ± 0.1303	0.8801 ± 0.0882	0.9337 ± 0.0095
$E_{29}$	0.9461 ± 0.0496	<b>0.9673 ± 0.0666</b>	0.9058 ± 0.1362	0.8575 ± 0.0834	0.9337 ± 0.0095
$E_{30}$	0.9556 ± 0.0454	<b>0.9691 ± 0.0537</b>	0.8907 ± 0.1369	0.8801 ± 0.0882	0.9337 ± 0.0095
$E_{31}$	0.9557 ± 0.0431	<b>0.9685 ± 0.0553</b>	0.8949 ± 0.1305	0.8801 ± 0.0882	0.9337 ± 0.0095
$E_{32}$	0.9557 ± 0.0431	0.9377 ± 0.1242	0.8949 ± 0.1305	0.8801 ± 0.0882	0.9337 ± 0.0095
$E_{33}$	0.9525 ± 0.0455	<b>0.9691 ± 0.0537</b>	0.8932 ± 0.1303	0.8801 ± 0.0882	0.9337 ± 0.0095

Tabela IV.10: Acurácia média  $\pm$  desvio-padrão dos *Ensembles Bagging* com opiniões implícitas do HCS-I e Máquinas do HCS-F no processo de validação cruzada com 10 conjuntos em cada *dataset*

### IV.3.4 Geração de Ensembles com Melhores Classificadores da Fase 2

Na Etapa 4 da segunda fase, análoga e etapa 4 da primeira fase, os melhores modelos de classificação gerados para cada dataset foram utilizados na formação de novos Ensembles, em que  $I=Votação$  e  $M$  o método de votação majoritária simples. O desempenho obtido pelos Ensembles com melhores classificadores da segunda fase pode ser observado no gráfico abaixo. As especificações e a consolidação dos resultados obtidos em cada dataset estão indicadas na tabela IV.11

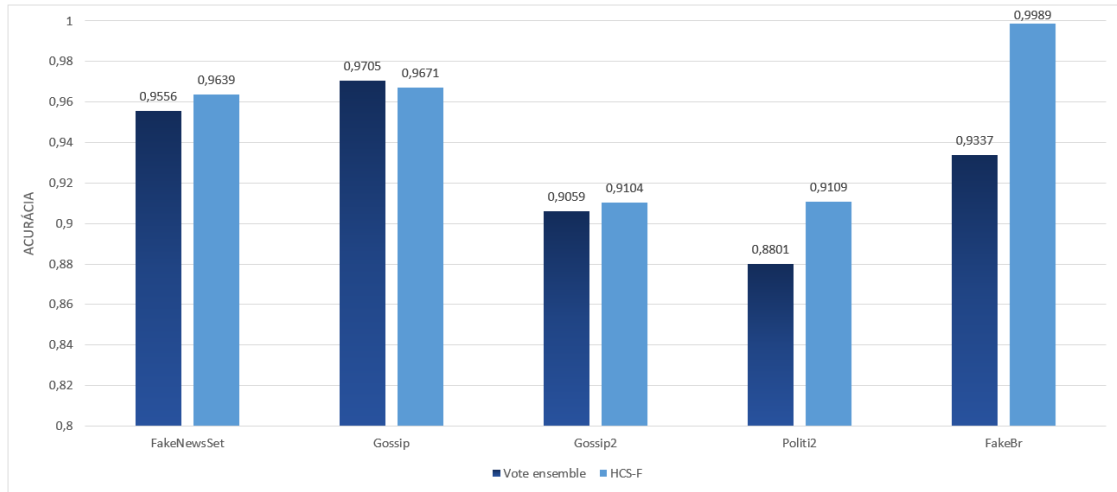


Figura IV.24: Acurácia dos Ensembles com Melhores Classificadores da Fase 2 em cada *Dataset*

<i>Ensembles</i>	Especificações	Dataset	Acurácia/Desvio Padrão	HCS-F
$E_{34}$	$C = \{E_{21}, E_{23}, E_{32}\}$	FakeNewsSet	$0.9556 \pm 0.0455$	<b><math>0.9639 \pm 0.0281</math></b>
$E_{35}$	$C = \{E_{21}, E_{26}, E_{27}\}$	Gossip	<b><math>0.9705 \pm 0.0502</math></b>	$0,9671 \pm 0,0094$
$E_{36}$	$C = \{E_{22}, E_{29}, E_{32}\}$	Gossip2	$0.9059 \pm 0.1363$	<b><math>0,9104 \pm 0,0249</math></b>
$E_{37}$	$C = \{E_{25}, E_{26}, E_{30}\}$	PolitiFact	$0.8801 \pm 0.0882$	<b><math>0,9109 \pm 0,0506</math></b>
$E_{38}$	$C = \{E_{20}, E_{22}, E_{23}\}$	FakeBr	$0.9337 \pm 0.0096$	<b><math>0,9989 \pm 0,0026</math></b>

Tabela IV.11: Especificações e resultados dos *Ensembles* com melhores classificadores - validação cruzada com 10 conjuntos em cada *dataset* (Fase 2)

### IV.3.5 Comparação e Análise de Resultados da Fase 2

Na etapa 5 da segunda fase, foi realizada a comparação dos resultados gerados nos experimentos. Foi possível observar que os *Ensembles* criados com opiniões implícitas do HCS-I e máquinas do HCS-F não obtiveram resultados superiores na maioria dos *datasets* executados, a pesar de apresentarem resultados muito próximos do desempenho do HCS-F nos *datasets* Gossip2 e FakeNewsSet. Contudo, o dataset Gossip foi o único que os ensembles desenvolvidos obtiveram resultados superiores ao HCS-F. Dos *Ensembles* executados no Gossip, apenas os que utilizaram os algoritmos KNN e LR como método para integração dos resultados, não apresentaram resultados superiores ao HCS-F. Em busca da variação dos modelos inicialmente desenvolvidos, a execução dos *Ensembles bagging* com opiniões

implícitas do HCS-I e máquinas do HCS-F e os *Ensembles* gerados com os melhores classificadores, obtiveram o desempenho semelhante aos *Ensembles* inicialmente desenvolvidos, obtendo acurácia superior ao HCS-F apenas no dataset Gossip e obtendo desempenho próximo ao HCS-F nos *datasets* Gossip2 e FakeNewsSet.

## Capítulo V Conclusão

### V.1 Considerações Finais

A abordagem computacional utilizada para detecção de Fake News denominada *Hybrid Crowd Signals* (HCS) apresenta bom potencial por não depender da contribuição voluntária dos usuários das redes sociais. Tal abordagem possui dois métodos: HCS-I, que utiliza apenas as opiniões implícitas dos usuários das redes sociais, e o método HCS-F, o qual combina as opiniões implícitas dos usuários e de diferentes máquinas de classificação. Apesar de possuir bons resultados, o método HCS-F, que possui melhor acurácia, conjuga as classificações de maneira ingênua. A partir desta premissa, este trabalho apresenta a hipótese de que utilizar *Ensemble* para combinar as classificações fornecidas pelas máquinas e as opiniões implícitas dos usuários sobre a notícia a ser detectada pode viabilizar a construção de métodos de detecção de fake news, baseados na abordagem Crowd Signals implícitos, mais robustos que o método HCS-F.

O presente trabalho teve como objetivo realizar experimentos com o intuito de avaliar se o uso de *Ensembles* como forma de integrar a opiniões implícitas de usuários e as opiniões explícitas de modelos de classificação gerados a partir de algoritmos de AM poderia superar o método HCS-F, para isso, os experimentos foram executados em duas fases: A primeira fase buscou avaliar se o uso de *Ensembles* como forma de integrar os pareceres de modelos de classificação gerados a partir de algoritmos de AM poderia superar o HCS-I, a fim de levantar evidências iniciais da validade da hipótese

Ao executar a primeira fase de experimentos, foi possível observar que os modelos ensembles *bagging* E9 e E11 obtiveram resultados superiores ao HCS-I nos *datasets* FakeNewsSet (com maior número de máquinas) e Gossip (com maior número de amostras) e o modelo Ensemble com melhores classificadores E15 também obteve desempenho superior ao HCS-I no dataset FakeNewsSet. Os experimentos da primeira fase produziram resultados em que, apesar de poucos *Ensembles* terem superado o HCS-I, a maioria apresentou resultados relativamente próximos aos do HCS-I, tais resultados foram considerados suficiente para validar a primeira fase e dar início à segunda fase de experimentos.

A segunda fase de experimentos consistiu na busca por modelos de melhor desempenho que o HCS-F, através da substituição da inferência bayesiana ingênua realizada neste método, pelos



ensembles desenvolvidos neste trabalho, com a finalidade de combinar as classificações fornecidas pelas opiniões implícitas dos usuários e as opiniões explícitas das máquinas presentes no método HCS-F. Ao executar os experimentos da segunda fase, foi possível constatar desempenho inferior dos ensembles nos datasets PolitiFact e FakeBr, desempenho inferior, porém bem próximo ao HCS-F nos dataset FakeNewsSet e Gossip2 e por fim desempenho superior em praticamente todos ensembles executados no dataset Gossip, exceto nos modelos E20 E25 e E32.

Analisando os resultados apresentados, é possível perceber que os métodos ensembles revelam um potencial promissor no âmbito da detecção de fake news. Apesar dos modelos desenvolvidos neste trabalho não terem superado o desempenho global do método HCS-F em todos os conjuntos de dados empregados, os ensembles gerados apresentaram um desempenho comparável aos métodos da abordagem HCS. Por outro lado, foi possível observar que os ensembles desenvolvidos neste estudo alcançaram resultados superiores aos métodos da abordagem HCS-I nos conjuntos de dados com maior quantidade de amostras (Gossip) e com maior número de máquinas (FakeNewsSet), além de apresentar resultados superiores ao método HCS-F no dataset Gossip. Embora os resultados não tenham demonstrado uma superioridade plena dos ensembles propostos, eles oferecem indícios significativos para validar a hipótese levanta de que a utilização de Ensembles para a combinação das opiniões explícitas das máquinas e as opiniões implícitas dos usuários sobre as notícias analisadas parece promissora e pode abrir caminho para a criação de métodos de detecção de Fake News mais robustos em relação ao método HCS-F.

## V.2 Trabalhos futuros

Para trabalhos futuros e aperfeiçoamento dos modelos gerados neste estudo, além da utilização de outras métricas de avaliação, as seguintes iniciativas poderão ser adotadas: (i) otimização dos valores dos hiperparâmetros dos algoritmos de classificação; (ii) busca pela melhoria do desempenho dos *Ensembles* ao fornecer ao método que implementa a estratégia de integração, além das saídas dos modelos de classificação, também as entradas recebidas por tais modelos; e (iii) incorporação dos *Ensembles* no método HCS-F.

## Referências

- Al Obaid, A., Khotanlou, H., Mansoorizadeh, M., and Zabihzadeh, D. Multimodal fake-news recognition using ensemble of deep learners. *Entropy*, 24(9):1242, 2022.
- Alghamdi, J., Lin, Y., and Luo, S. A comparative study of machine learning and deep learning techniques for fake news detection. *Information*, 13(12):576, 2022.
- Ali, A. M., Ghaleb, F. A., Al-Rimy, B. A. S., Alsolami, F. J., and Khan, A. I. Deep ensemble fake news detection model using sequential deep learning technique. *Sensors*, 22(18):6970, 2022.
- Ayyub, K., Iqbal, S., Nisar, M. W., Munir, E. U., Alarfaj, F. K., and Almusallam, N. A feature-based approach for sentiment quantification using machine learning. *Electronics (Switzerland)*, 11(6), 2022.
- Biradar, S., Saumya, S., and Chauhan, A. Combating the infodemic: Covid-19 induced fake news recognition in social media networks. *Complex and Intelligent Systems*, 2022.
- Breiman, L. Bagging predictors. *Mach Learn*, 24(2):123–140, 1996.
- Brunette, E. S., Flemmer, R. C., and Flemmer, C. L. A review of artificial intelligence. In *2009 4th International Conference on Autonomous Robots and Agents*, pages 385–392. Ieee, 2009.
- Castillo, J. M., Fadera, K. D. F., Ladao, A. A. A., Go, J. G., Tamayo, M. B., and Octaviano, M. V. Fake news detection on english news article’s title. In *Proceedings - 2021 1st International Conference in Information and Computing Research, iCORE 2021*, pages 151–156, 2021.
- Chabukswar, A., Sushmitha, J., Sandhya, G., Likhith, S., Shenoy, P. D., and Venugopal, K. Fake news detection on covid-19 tweeter data using machine learning models. In *2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)*, pages 1–7. IEEE, 2022.
- Das, S. D., Basak, A., and Dutta, S. A heuristic-driven uncertainty based ensemble framework for fake news detection in tweets and news articles. *Neurocomputing*, 491:607–620, 2022.
- de Souza, M. P., da Silva, F. R. M., Freire, P. M. S., and Goldschmidt, R. R. A linguistic-based method that combines polarity, emotion and grammatical characteristics to detect fake news in portuguese. WebMedia, 2020.

- Freire, P. and Goldschmidt, R. Fake news detection on social media via implicit crowd signals. pages 521–524, 2019.
- Ganesh, P., Priya, L., and Nandakumar, R. Fake news detection - a comparative study of advanced ensemble approaches. In *Proceedings of the 5th International Conference on Trends in Electronics and Informatics, ICOEI 2021*, pages 1003–1008, 2021.
- Gethsia, P. and Juliet, S. An enhanced approach for fake news detection using ensemble techniques. In *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 1957–1962. IEEE, 2023.
- Gravanis, G., Vakali, A., Diamantaras, K., and Karadais, P. Behind the cues: A benchmarking study for fake news detection. *Expert Systems with Applications*, 128:201–213, 2019.
- Guo, B., Ding, Y., Yao, L., Liang, Y., and Yu, Z. The future of false information detection on social media: New perspectives and trends. *ACM Computing Surveys (CSUR)*, 53(4):1–36, 2020.
- Hansrajh, A., Adeliyi, T. T., and Wing, J. Detection of online fake news using blending ensemble learning. *Scientific Programming*, 2021, 2021.
- Imbwaga, J. L., Chittaragi, N., and Koolagudi, S. Fake news detection using machine learning algorithms. In *Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing*, pages 271–275, 2022.
- Jlifi, B., Sakrani, C., and Duvallet, C. Towards a soft three-level voting model (soft t-lvm) for fake news detection. *Journal of Intelligent Information Systems*, pages 1–21, 2022.
- Kaliyar, R. K., Goswami, A., and Narang, P. Multiclass fake news detection using ensemble machine learning. In *Proceedings of the 2019 IEEE 9th International Conference on Advanced Computing, IACC 2019*, pages 103–107, 2019.
- Khan, M. Z. and Alhazmi, O. H. Study and analysis of unreliable news based on content acquired using ensemble learning (prevalence of fake news on social media). *International Journal of System Assurance Engineering and Management*, 11:145–153, 2020.
- Liu, Y. and Zhou, Y. An ensemble learning approach for covid-19 fact verification. In *2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pages 383–387. IEEE, 2022.
- Lofi, C. and Maarry, K. E. Design patterns for hybrid algorithmic-crowdsourcing workflows. In *2014 IEEE 16th Conference on Business Informatics*, volume 1, pages 1–8, 2014.

- Ludermir, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avancados*, 35(101):85–94, 2021.
- Meel, P. and Vishwakarma, D. K. Han, image captioning, and forensics ensemble multimodal fake news detection. *Information Sciences*, 567:23–41, 2021.
- Mitchell, T. *Machine Learning*, volume 1. McGraw-Hill, 1997.
- Moraes, M. P., de Oliveira Sampaio, J., and Charles, A. C. Data mining applied in fake news classification through textual patterns. WebMedia, 2019.
- Panda, R. and Kumari, S. Analysis of deep ensemble transformer model for fake news detection. In *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, pages 1–5. IEEE, 2022.
- Priya, A. and Kumar, A. Deep ensemble approach for covid-19 fake news detection from social media. In *Proceedings of the 8th International Conference on Signal Processing and Integrated Networks, SPIN 2021*, pages 396–401, 2021.
- Rahman, J. U., Ali, S., Khan, A., and Fayaz, M. A deep ensemble model for news classification on social media. In *2021 International Conference on Cyber Warfare and Security, ICCWS 2021 - Proceedings*, pages 57–62, 2021.
- Ramkissoon, A. N. and Goodridge, W. Legitimacy: An ensemble learning model for credibility based fake news detection. In *IEEE International Conference on Data Mining Workshops, ICDMW*, volume 2021-December, pages 254–261, 2021.
- Rezaei, S., Kahani, M., Behkamal, B., and Jalayer, A. Early multi-class ensemble-based fake news detection using content features. *Social Network Analysis and Mining*, 13(1):16, 2022.
- Russel, S. and Norvig, P. *Artificial Intelligence: A Modern Approach*, volume 3. Prentice Hall, 2009.
- Sandrilla, R. and Devi, M. S. Fnu-bicnn: Fake news and fake url detection using bi-cnn. *International Journal of Advanced Computer Science and Applications*, 13(2):477–488, 2022.
- Sangamnerkar, S., Srinivasan, R., Christhuraraj, M. R., and Sukumaran, R. An ensemble technique to detect fabricated news article using machine learning and natural language processing techniques. In *2020 International Conference for Emerging Technology, INCET 2020*, 2020.
- Sharmin, S., Saha, S., Hoque, T., and Nadib, K. A. Interaction based credibility analysis of news on facebook using machine learning methodologies. In *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 465–471. IEEE, 2022.

- Souza Freire, P. M., Matias da Silva, F. R., and Goldschmidt, R. R. Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning. *Expert Systems with Applications*, 183, 2021.
- Tschiatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., and Krause, A. Fake news detection in social networks via crowd signals. In *Companion proceedings of the the web conference 2018*, pages 517–524, 2018.
- Turing, A. M. I.—Computing Machinery and Intelligence. *Mind*, LIX(236):433–460, 1950.
- Varshney, D. and Vishwakarma, D. K. An automated multi-web platform voting framework to predict misleading information proliferated during covid-19 outbreak using ensemble method. *Data & Knowledge Engineering*, 143:102103, 2023.
- Vasist, P. N. and Sebastian, M. Tackling the infodemic during a pandemic: A comparative study on algorithms to deal with thematically heterogeneous fake news. *International Journal of Information Management Data Insights*, 2(2):100133, 2022.
- Wynne, H. E. and Swe, K. T. Fake news detection in social media using two-layers ensemble model. In *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, pages 411–414. IEEE, 2022.
- Zhan, K., Li, Y., Osmani, R., Wang, X., and Cao, B. Data exploration and classification of news article reliability: Deep learning study. *JMIR infodemiology*, 2(2):e38839, 2022.
- Zhang, C. and Ma, Y. *Ensemble Machine Learning: Methods and Applications*. Springer, 2012.