

UNIVERSIDADE FEDERAL DO MARANHÃO

CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE

Marcus Vinicius de Sousa Lopes

*Aplicação de Classificadores para Determinação de
Conformidade de Biodiesel*

São Luís

2017

Marcus Vinicius de Sousa Lopes

*Aplicação de Classificadores para Determinação de
Conformidade de Biodiesel*

Tese apresentada ao Programa de Pós-Graduação
em Engenharia de Eletricidade da UFMA, como
parte dos requisitos necessários para obtenção do
grau de DOUTOR em Engenharia de Eletricidade.

Orientador: Prof. Dr. Allan Kardec Duailibe Barros Filho

Universidade Federal do Maranhão

São Luís

2017

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Núcleo Integrado de Bibliotecas/UFMA

Lopes, Marcus Vinicius de Sousa.

Aplicação de Classificadores para Determinação de Conformidade de Biodiesel / Marcus Vinicius de Sousa Lopes. - 2017.

82 f.

Orientador(a): Allan Kardec Duailibe Barros Filho.

Tese (Doutorado) - Programa de Pós-graduação em Engenharia de Eletricidade/ccet, Universidade Federal do Maranhão, Laboratório PIB 3/CCET/UFMA, 2017.

1. Árvore de Decisões. 2. Biodiesel. 3. K-vizinhos próximos. 4. Máquina de vetor de suporte. 5. Parâmetros de qualidade. I. Barros Filho, Allan Kardec Duailibe. II. Título.

Marcus Vinicius de Sousa Lopes

*Aplicação de Classificadores para Determinação de
Conformidade de Biodiesel*

Tese apresentada ao Programa de Pós-Graduação
em Engenharia de Eletricidade da UFMA, como
parte dos requisitos necessários para obtenção do
grau de DOUTOR em Engenharia de Eletricidade.

Aprovado em 26 de julho de 2017

BANCA EXAMINADORA

Prof. Dr. Allan Kardec Duailibe Barros Filho
Universidade Federal do Maranhão

Prof. Dr. Ewaldo Eder Carvalho Santana
Universidade Federal do Maranhão

Prof. Dr. João Viana da Fonseca Neto
Universidade Federal do Maranhão

Profa. Dra. Kiany Sirley Brandão Cavalcante
Instituto Federal do Maranhão

Profa. Dr. Fausto Lucena de Oliveira
Universidade CEUMA

À Ruth Maria, Patricia e Maria Victoria.

Agradecimentos

Ao meu orientador Prof. Allan Kardec, pela confiança e pela oportunidade concedida.

Ao Prof. Ewaldo Santana, por toda a confiança creditada e pelos ensinamentos sobre objetividade.

À Profa. Aldalea Marques e Prof. Edmar Marques pela contribuição valorosa na área de biodiesel.

Aos colegas do Laboratório de Processamento da Informação Biológica (PIB), em especial a Luis Cláudio Silva que esteve comigo nessa caminhada sempre compartilhando conhecimento e Christian Diniz pela grande ajuda e sempre com um conselho de serenidade.

Aos colegas do Laboratório de Análise e Pesquisa e Química Analítica de Petróleo (LAPQAP), em especial a Alex Barradas Filho e Isabelle Viegas pela parceria de grande valia neste trabalho.

Ao Programa de Pós-Graduação em Engenharia de Eletricidade da UFMA, apoio administrativo e todos os professores que me auxiliaram nessa jornada, especialmente à Profa. Maria da Guia pelo carinho e confiança depositada.

À CAPES pelo financiamento deste trabalho.

Resumo

LOPES, M. V. de S. **Aplicação de Classificadores para Determinação de Conformidade de Biodiesel**. 2017. 82 f. Tese - Programa de Pós-Graduação em Engenharia de Eletricidade, Universidade Federal do Maranhão, São Luís, 2017.

A demanda crescente por fontes de energia renováveis e como alternativa aos combustíveis fósseis tornam o biodiesel como uma das principais alternativas para substituição dos derivados do petróleo. O controle da qualidade do biodiesel durante processo de produção e distribuição é extremamente importante para garantir um combustível com qualidade confiável e com desempenho satisfatório para o usuário final. O biodiesel é caracterizado pela medição de determinadas propriedades de acordo com normas internacionais. A utilização de métodos de aprendizagem de máquina para a caracterização do biodiesel permite economia de tempo e dinheiro. Neste trabalho é mostrado que para a determinação da conformidade de um biodiesel os classificadores SVM, KNN e Árvore de decisões apresentam melhores resultados que os métodos de predição de trabalhos anteriores. Para as propriedades de viscosidade densidade, índice de iodo e estabilidade oxidativa (RANP 45/2014, EN14214:2014 e ASTM D6751-15) os classificadores KNN e Árvore de decisões apresentaram-se como melhores opções. Estes resultados mostram que os classificadores podem ser aplicados de forma prática visando economia de tempo, recursos financeiros e humanos.

Palavras-chave: Biodiesel, Parâmetros de qualidade, Máquina de vetor de suporte, K-vizinhos próximos, Árvore de Decisões.

Abstract

LOPES, M. V. de S. **Attesting compliance of biodiesel quality using classification methods**. 2017. 82 f. Tese - Programa de Pós-Graduação em Engenharia de Eletricidade, Universidade Federal do Maranhão, São Luís, 2017.

The growing demand for energy and the limitations of oil reserves have led to the search for renewable and sustainable energy sources to replace, even partially, fossil fuels. Biodiesel has become in last decades the main alternative to petroleum diesel. Its quality is evaluated by given parameters and specifications which vary according to country or region like, for example, in Europe (EN 14214), US (ASTM D6751) and Brazil (RANP 45/2014), among others. Some of these parameters are intrinsically related to the composition of fatty acid methyl esters (FAMEs) of biodiesel, such as viscosity, density, oxidative stability and iodine value, which allows to relate the behavior of these properties with the size of the carbon chain and the presence of unsaturation in the molecules. In the present work four methods for direct classification (support vector machine, K-nearest neighbors, decision tree classifier and artificial neural networks) were optimized and compared to classify biodiesel samples according to their compliance to viscosity, density, oxidative stability and iodine value, having as input the composition of fatty acid methyl esters, since those parameters are intrinsically related to composition of biodiesel. The classifications were carried out under the specifications of standards EN 14214, ASTM D6751 and RANP 45/2014. A comparison between these methods of direct classification and empirical equations (indirect classification) distinguished positively the direct classification methods in the problem addressed, especially when the biodiesel samples have properties values very close to the limits of the considered specifications.

Keywords: Biodiesel, Quality Parameters, Support Vector Machine, K-Nearest Neighbors, Decision Tree Classifier.

“O uso de óleos vegetais como combustível pode parecer insignificante hoje em dia. Mas com o tempo, esses produtos irão se tornar tão importantes quanto o querosene e o carvão são atualmente.”

Rudolph Diesel, 1911

Sumário

1	Introdução	16
1.1	Trabalhos Relacionados	17
1.2	Objetivos	22
1.2.1	Objetivo geral	22
1.2.2	Objetivos específicos	22
1.3	Contribuições	22
1.4	Organização do Trabalho	22
2	Fundamentação Teórica	24
2.1	Biodiesel	24
2.1.1	Benefícios do biodiesel e conformidade com os padrões industriais	25
2.1.2	Propriedades	26
2.1.3	Padrões para Caracterização do Biodiesel	28
2.1.3.1	EN 14214	29
2.1.3.2	ASTM D6751	29
2.1.3.3	Resolução ANP 45/2014	30
2.1.4	Controle de qualidade em diferentes escalas de produção	32
2.1.5	Viabilidade Econômica do Biodiesel	33
2.2	Classificadores	35
2.2.1	Máquina de Vetor de Suporte	35
2.2.2	K-nearest Neighbor	40
2.2.3	Árvore de Decisões	40
2.3	Bases desbalanceadas	43
2.4	Misturas	44
2.4.1	Mistura de dois componentes	46
2.4.2	Mistura de três componentes	46
3	Metodologia Proposta	47
3.1	Base de dados	47
3.2	Normalizar/Padronizar/Escalonar	50
3.3	Modelos dos classificadores	52

3.4	Medidas de desempenho	54
4	Resultados e Discussões	58
4.1	Resultados do algoritmo SMOTE	58
4.2	Resultados dos modelos de classificadores	63
4.3	Resultados dos modelos preditivos	66
4.4	Software	68
5	Conclusão, Publicações e Trabalhos Futuros	70
5.1	Conclusão	70
5.2	Trabalhos Publicados	71
5.3	Trabalhos Futuros	71
	Referências Bibliográficas	72
	Apêndice A Software	80

Lista de Abreviaturas

ACC	Acurácia
ASTM	Sociedade Americana para Ensaio de Materiais (<i>American Society of Testing and Materials</i>)
ANN	Redes Neurais Artificiais (<i>Artificial Neural Network</i>)
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
C8:0	Ácido caprílico
C10:0	Ácido cáprico
C12:0	Ácido láurico
C14:0	Ácido mirístico
C16:0	Ácido palmítico
C18:0	Ácido esteárico
C18:1	Ácido oleico
C18:2	Ácido linoleico
C18:3	Ácido linolênico
C20:0	Ácido araquídico
C20:1	Ácido gadoleico
C22:1	Ácido erúcido
C18:1 OH	Ácido ricinoleico
CEN	Comitê Europeu de Normalização
CCPP	Ponto de entupimento do filtro a frio (<i>cold-filter plugging point</i>)
FAME	Éster metílico de ácidos graxos (<i>Fatty acid methyl ester</i>)
LAPQAP	Laboratório de Análise e Pesquisa e Química Analítica de Petróleo
NIR	Infra-vermelho próximo (<i>Near infrared</i>)
MLR	Regressões lineares múltiplas (<i>Multiple linear regression</i>)
OVEG	Programa de Óleos Vegetais
PCR	Regressão de componentes principais (<i>Principal component regression</i>)
PIB	Laboratório de Processamento da Informação Biológica
PLS	Regressão de mínimos quadrados parciais (<i>Partial least squares regression</i>)
RANP 45/2014	Resolução ANP 45/2014
RDA	Análise discriminante regularizada (<i>Regularized discriminant analysis</i>)
PLS-DA	Análise de discriminante com método de mínimos quadrados parciais

	<i>(Partial least squares discriminant analysis)</i>
KNN	K-vizinho próximo (<i>K-nearest neighbors</i>)
SVM	Máquinas de vetor de suporte (<i>Support vector machines</i>)
HCA	Modelos baseados em análises de agrupamentos hierárquicos (<i>Hierarchical cluster analysis</i>)
PCA	Análise de componentes principais (<i>Principal component analysis</i>)
SENS	Sensibilidade
SIMCA	Modelagem independente e flexível por analogia de classe (<i>Soft independent modeling of class analogy</i>)
SMOTE	Técnica de sobreamostragem de minoria sintética (<i>Synthetic minority over-sampling technique</i>)
SPEN	Especificidade
TREE	Árvore de decisões
UFMA	Universidade Federal do Maranhão

Lista de Figuras

2.1	Reação de transesterificação de triglicerídeos.	25
2.2	Matérias-primas utilizadas para produção de biodiesel no Brasil (Abril/2017)	26
2.3	Evolução anual da produção, demanda compulsória e capacidade nominal.	34
2.4	Produção, demanda compulsória e capacidade nominal por região	35
2.5	Conjunto de treinamento contendo duas classes com vetores de suporte e hiperplano ótimo	36
2.6	Conjunto de treinamento contendo duas classes com vetores de suporte e hiperplano ótimo	37
2.7	Árvore de decisão para um problema de classificação de mamífero.	41
2.8	Partição binária e multicaminhos.	42
2.9	Regiões do espaço simplex para misturas de 2, 3 e 4 componentes.	45
3.1	Frequência de cada éster para as amostras de biodiesel da base de dados utilizada.	48
3.2	Projeção das 98 amostras para visualização em 2 dimensões utilizando PCA agrupadas por matéria-prima.	49
3.3	Projeção das 98 amostras para visualização em 3 dimensões utilizando PCA para as propriedades de viscosidade, índice de iodo, estabilidade oxidativa e massa específica para as especificações EN 14214, ASTM D6751 e RANP 45/2014	51
3.4	Fluxograma com as principais partes da metodologia utilizada no trabalho.	57
4.1	Amostras utilizadas para teste da propriedade: viscosidade (a) RANP 45/2014 (c) EN 14214:2014 e (e) ASTM D6751-15; estabilidade oxidativa (b) RANP 45/2014 e EN 14214:2014, (d) ASTM D6751-15; índice de iodo (f) (EN 14214:2014) e massa específica (g) RANP 45/2014 e (h) EN 14214:2014. . .	59
4.2	Médias dos resultados obtidos para SVM, KNN, TREE e ANN para o conjunto de amostras externo: comparação entre os dados originais e com a aplicação do algoritmo SMOTE	61

4.3	Comparação do desempenho entre os métodos SVM, KNN, TREE e ANN para classificação de amostras de biodiesel do conjunto de validação externo, para as propriedades de viscosidade, índice de iodo, estabilidade oxidativa e massa específica	63
4.4	Comparação entre métodos de classificação direta (presente trabalho) e classificação indireta (equações da literatura) aplicada ao conjunto de validação externa para a classificação de biodiesel de acordo com as propriedades de viscosidade, índice de iodo, estabilidade oxidativa e massa específica	67
4.5	Resultados de classificação para as propriedade de massa específica (a), viscosidade (b), estabilidade oxidativa(c) e índice de iodo (d) do conjunto de teste de acordo com o tipo de matéria-prima utilizado.	69
A.1	Telas de instalação.	80
A.2	Classificação para uma amostra.	81
A.3	Classificação para duas amostras (mistura binária).	81
A.4	Classificação para três amostras (mistura ternária).	82

Lista de Tabelas

2.1	Comparação entre as propriedades especificadas pelas normas RANP 45/2014, ASTM D6751, EN 14214 e CAN/CGSB-3524-2011.	31
2.2	Exemplo da geração de dados utilizando SMOTE	44
3.1	Amostras de biodiesel utilizadas e especificações adotadas para a verificação dos parâmetros de qualidade	48
3.2	Parâmetros para a construção dos modelos de classificador SVM	52
3.3	Parâmetros para construção dos modelos de classificação KNN	52
3.4	Parâmetros de treinamento das Redes Neurais	53
3.5	Resultados da otimização das redes neurais	54
3.6	Modelos lineares utilizados para predição das propriedades de viscosidade, índice de iodo, estabilidade oxidativa e massa específica	56
4.1	Número de amostras para cada classe de acordo com as especificações das normas EN 14214, ASTM D6751-15 e RANP 45/2014 e o número de amostras sintéticas produzidas com o algoritmo SMOTE (entre parêntesis), excluindo as 6 amostras utilizadas para validação externa	58
4.2	Resultados obtidos para SVM, KNN, TREE e ANN para classificação de viscosidade, índice de iodo, estabilidade oxidativa e massa específica para as normas EN 14214, ASTM D6751-15 e RANP 45/2014	64

Capítulo 1

Introdução

Nos últimos anos o interesse do biodiesel como uma substituição aos combustíveis fósseis tem aumentado significativamente. Fatores como redução de emissão de gases de efeito estufa e mudanças do clima global, busca por fontes de energia renováveis tem influenciado esse interesse (HOEKMAN et al., 2012). O biodiesel é uma das principais alternativas para a substituição dos derivados do petróleo como uma fonte de energia alternativa.

A Resolução da ANP nº 45/2014 define o biodiesel como: “combustível composto de alquil ésteres de ácidos graxos de cadeia longa, derivados de óleos vegetais ou de gorduras animais” (ANP, 2014). O biodiesel pode ser utilizado diretamente como combustível o que requer algumas modificações nos motores ou misturado com diesel comum, necessitando pouca ou nenhuma modificação (STOYTCHIEVA et al., 2011; SARIN, 2012).

A Lei nº 11.097/2005 introduziu o biodiesel na matriz energética brasileira, tratando a fiscalização e regulamentação do mercado (BRASIL, 2005). Em 2008 se torna obrigatória a adição de 2% de biodiesel em todo o diesel comercializado no Brasil, em 2013 esse valor é aumentado para 5% e em 1º de novembro de 2014 a adição obrigatória passou a ser de 7%.

A Lei nº 13.263/2016 determinou um novo cronograma de aumento do teor de biodiesel a partir de 2017 (BRASIL, 2016). A última mudança foi em 1º de março de 2017 onde o percentual obrigatório de biodiesel passou para 8%. Seguindo o cronograma, as composições deverão ser elevadas para 9% e 10%, respectivamente, a partir de 1º de março de 2018 e 1º de março de 2019. Com esses novos percentuais de adição do biodiesel o Brasil passa para uma posição de destaque no mercado internacional do biocombustível.

O Brasil é o segundo maior produtor de biocombustíveis no mundo (22,5%), seguindo os Estados Unidos (43,5%) e deixando para trás produtores europeus tradicionais

(BP, 2017). Com o novo percentual de 8% o Brasil prevê uma redução das importações de óleo diesel e busca favorecer a agricultura familiar e o agronegócio brasileiro incentivando o cultivo de matéria prima para produção de biodiesel.

A maior parte do biodiesel produzido é proveniente de óleos comestíveis com metanol, na presença de um catalisador alcalino. O produto obtido tem baixa viscosidade e é um biocombustível (éster metil de ácido graxo). Os óleos vegetais precisam ser modificados para a que suas propriedades estejam próximas do diesel mineral. A principal modificação é reduzir a viscosidade. Diversos esforços tem sido feitos para que o desempenho do biodiesel se aproxime dos combustíveis fósseis (STOYTCHEVA et al., 2011; SARIN, 2012).

Óleos refinados de oleaginosas de baixa acidez podem facilmente ser convertidos em biodiesel, contudo a sua utilização aumenta significativamente os custos de produção, resultando em um biodiesel não competitivo com o diesel proveniente de petróleo. Presumindo que com o aumento do mercado e tecnologia utilizada é melhorada, o custo irá então cair (STOYTCHEVA et al., 2011). Com tudo isso os custos da matéria prima podem chegar facilmente a 75% dos custos do biodiesel. A viabilidade de qualquer matéria prima para a produção de biodiesel depende do clima, geografia, condições de solo e práticas agrícolas de cada região (ATABANI et al., 2012).

A exploração de culturas não-comestíveis com baixo custo agrícola e a utilização de subprodutos na produção do biodiesel podem reduzir significativamente os custos (ATABANI et al., 2012; ATABANI et al., 2013). A utilização de misturas de diferentes matérias-primas permite então a produção do biodiesel economicamente viável para cada região.

1.1 Trabalhos Relacionados

Ayeter, Sunnu e Parbey (2015) estudaram as características do biodiesel produzido a partir dos óleos de semente de dendê, pinhão-manso e coco bem como a determinação das misturas nas proporções de 100%, 75%, 50% e 25% que resultavam nas melhores propriedades físico-químicas e melhor desempenho para o motor diesel. Os melhores valores para emissões de escape e consumo de combustível foram obtidos com 75% de coco e 25% de semente de dendê. Foi observado ainda que a mistura de 75% de pinhão-manso e 25% de coco produziu melhor desempenho no motor que o diesel convencional.

Sanjid et al. (2014) investigaram as propriedades de potência, emissão de gases e ruído, combinando mistura de dendê e pinhão-manso com diesel. As emissões de CO para

as misturas do tipo B5 e B10 foram 9,53% e 20,49% mais baixas que o diesel puro. O ruído do motor também foi reduzido em 2,5% e 5% respectivamente, quando comparado ao diesel, devido às características de lubrificação e amortecimento.

Al-Hamamre e Al-Salaymeh (2014) estudaram as propriedades de viscosidade, densidade, ponto de fulgor e poder calorífico do óleo de jojoba misturado a biodiesel e diesel. As relações entre densidade e temperatura foram expressas por meio de equações de 3º grau com máximo erro absoluto de 0,9921 para as misturas de biodiesel-diesel. Para misturas de jojoba com biodiesel e jojoba com diesel a viscosidade aumenta com o aumento a porcentagem de óleo de jojoba. O ponto de fulgor das diferentes misturas foi representado por uma equação de 6º grau enquanto que o poder calorífico variou linearmente.

Atabani et al. (2013) fez um estudo das propriedades óleos comestíveis e não comestíveis disponíveis no sudoeste asiático: *Jatropha curcas* (pinhão-manso), *Calophyllum inophyllum*, *Sterculia foetida*, *Moringa oleifera* (acácia-branca), *Croton megalocarpus*, Patchouli, *Elaeis guineensis* (dendê), *Cocos nucifera* (coco), *Brassica napus* (canola) e *Glycine Max* (soja). Foram estudadas as propriedades viscosidade, ponto de congelamento, ponto de fluidez e ponto de entupimento de filtro a frio. Discutiu o conceito de mistura biodiesel-biodiesel como forma de melhorar as propriedades do biodiesel produzido. As propriedades das misturas foram estimadas utilizando o método de aproximação de curva polinomial e concluiu que a adição de biodiesel de óleo não comestível na mistura biodiesel-biodiesel é uma alternativa para melhorar as propriedades do biodiesel final.

Hong, Lee e Lee (2015) estudaram o biodiesel produzido da mistura de óleo de canola com banha de porco. A relação molar álcool óleo depende da mistura canola e banha de porco. Também encontraram que a razão molar álcool/óleo ótima de óleo de canola e banha de porco foi de 3:7 utilizando metanol e 1:9 quando o etanol era utilizado na produção do biodiesel.

Yusup e Khan (2010) introduziram a mistura de óleo de dendê com óleo de semente de seringueira com o objetivo de diminuir a utilização do óleo comestível e investigaram as propriedades de densidade, viscosidade, poder calorífico, índice de acidez, teor de ácidos graxos livres, índice refrativo, teor de mono, di e triglicerídeos e enxofre.

Moser (2008) mostrou que o éster metílico de óleo de soja pode ser melhorado satisfazendo a especificação EN 14214 para as propriedades de índice de iodo e período de indução ao ser misturado com canola, dendê e girassol. O dendê teve o ponto de entupimento de filtro a frio melhorado quando misturado com canola, soja e girassol. Foram descobertas relações estatisticamente significativas entre período de indução e índice de iodo, período de indução e teor de óleo de girassol, período de indução e ponto de entupimento de filtro a frio, ponto de entupimento de filtro a frio e índice de iodo e ponto de

entupimento de filtro a frio e teor de óleo de girassol.

Outros autores verificaram as propriedades de misturas biodiesel-diesel para diferentes matérias-primas: camelina (ÖZÇELİK; AYDOĞAN; ACAROĞLU, 2015), dendê (BENJUMEA; AGUDELO; AGUDELO, 2008; MOFIJUR et al., 2014; ATABANI et al., 2014; SARIN et al., 2007), pinhão-manso (SARIN et al., 2007; SILITONGA et al., 2013), nogueira-de-iguape, *Sterculia foetida*, *Calophyllum inophyllum*, mafumeira, pangium, semente de seringueira e cerbera (SILITONGA et al., 2013); moringa (MOFIJUR et al., 2014; ATABANI et al., 2014); algodão (ATABANI et al., 2014), coco (ATABANI et al., 2014); soja (QI; LEE, 2014; CANDEIA et al., 2009; SANTOS et al., 2011; DANTAS et al., 2011b), girassol (TAVARES et al., 2011; SANTOS et al., 2011), milho (DANTAS et al., 2011a), mamona (DANTAS et al., 2011b; VALENTE et al., 2011) e óleo de fritura (VALENTE et al., 2011).

Oghenejoboh e Umukoro (2011) relataram que a mistura de biodiesel produzido a partir de dendê, semente de dendê, pinhão-manso e semente da árvore-da-borracha com diesel resultam em aumento de poder calorífico, diminuição de densidade, ponto de congelamento, viscosidade e ponto de fulgor.

Focando em abordagens não-lineares, Barradas Filho et al. (2015) otimizaram modelos de redes neurais artificiais (ANN) para prever viscosidade, índice de iodo e período de indução (estabilidade oxidativa) do biodiesel pela sua composição éster metílico de ácidos graxos (FAME, do inglês: *fatty acid methyl ester*). Utilizando espectroscopia de infravermelho próximo (NIR, do inglês: *near infrared*), Balabin, Lomakina e Safieva (2011) compararam regressões lineares múltiplas (MLR, do inglês: *multiple linear regression*), regressão de componentes principais (PCR, do inglês: *principal component regression*), regressão de mínimos quadrados parciais (PLS, do inglês: *partial least squares regression*), as versões polinomial e *Spline*-PLS e ainda ANN para a predição de massa específica, viscosidade cinemática, teor de metanol e água.

Além dos métodos de regressão, os métodos de classificação também têm sido aplicados a biodiesel, entretanto de forma mais restrita. Nos trabalhos de Balabin e Safieva (2011) e de Veras et al. (2010), dados de NIR foram utilizados para classificação de biodiesel de acordo com sua matéria-prima. Balabin e Safieva utilizaram análise discriminante regularizada (RDA, do inglês: *regularized discriminant analysis*), análise de discriminante com método de mínimos quadrados parciais (PLS-DA, do inglês: *partial least squares discriminant analysis*), K-vizinho próximo (KNN, do inglês: *K-nearest neighbors*), e máquinas de vetor de suporte (SVM, do inglês: *support vector machines*), enquanto que Veras et al. utilizaram modelos baseados em análises de agrupamentos hierárquicos, (HCA, do inglês: *hierarchical cluster analysis*), análise de componentes principais (PCA, do inglês: *principal component analysis*) e modelagem independente e flexível por analogia de classe

(SIMCA, do inglês: *soft independent modeling of class analogy*).

Chavarria-Hernandez e Pacheco-Catalán (2014) apresentaram três modelos empíricos para predição da viscosidade em uma ampla faixa de temperatura e para uma ampla faixa de cadeias de hidrocarbonetos. Também com modelos empíricos, Ramírez-Verduzco, Rodríguez-Rodríguez e Jaramillo-Jacob (2012) apresentaram a caracterização de duas amostras de biodiesel de sebo e soja, desenvolvendo equações para estimação de quatro propriedades físicas: número de cetanos, viscosidade cinemática, massa específica e poder calorífico.

Hong, Jeon e Lee (2014) fizeram a predição das propriedades de poder calorífico, estabilidade oxidativa viscosidade cinemática analisando as características de vários alquil ésteres de ácidos graxos, obtendo como faixa de erro médio absoluto 0,14% a 1,42%, 0,4% a 7,5% e de 0,3% a 4,3% respectivamente, garantindo que várias propriedades de biodiesel podem ser preditas a partir de alquil ésteres de ácidos graxos.

Yang et al. (2002) testaram diferentes arquiteturas de redes neurais com a finalidade de estabelecer um modelo para correlacionar número de cetano e massa específica com a composição química. Os resultados foram comparados com regressão linear, obtendo melhores resultados para número de cetano, porém para a massa específica os resultados obtidos foram similares.

Também utilizando redes neurais Meng, Jia e Wang (2014) desenvolveram um modelo para predição da viscosidade em 313 K apresentando erro quadrático médio de 0,0099 a partir de 105 amostras coletadas na literatura. A ANN utiliza somente frações de massas de 19 ésteres metílicos de ácidos graxos como entradas, o que evita a necessidade da viscosidade dos FAMES individuais necessários para métodos preditivos usando equações de mistura.

Alves, Henriques e Poppi (2012) demonstraram a aplicação de máquina de vetor de suporte para regressão (SVR, do inglês: *support vector regression*) para determinação de parâmetros de qualidade de óleo diesel para um sistema de otimização de misturas em tempo real em uma refinaria de petróleo.

Shi et al. (2013) utilizaram SVR para otimização da eficiência de um motor utilizando éster metílico de ácido graxo de soja, afirmando que o processo de conversão vapor-calor-potência do motor pode ser precisamente simulado utilizando máquinas de vetor de suporte.

Alves e Poppi (2013) verificaram o potencial de máquinas de vetor de suporte para regressão utilizando espectroscopia de infra-vermelho próximo para desenvolver modelos para determinação do conteúdo em misturas de combustível.

Balabin, Safieva e Lomakina (2011) avaliaram a eficiência de diferentes métodos para classificações óleos por matéria-prima (sintético, semi-sintético e mineral), viscosidade em baixa temperatura (SAE 0W, SAE 5W, SAE 10W and SAE 15W) e viscosidade em alta temperatura (SAE 20W, SAE 30W, SAE 40W and SAE 50W). Rede neural probabilística (PNN, do inglês: *probabilistic neural network*) mostrou-se como o método mais eficiente, tendo SVM também apresentado um bom desempenho de classificação. Técnicas mais simples como KNN não se mostraram adequadas.

Poucos trabalhos na literatura utilizam métodos de classificação focados em parâmetros de qualidade de combustíveis. Pontes et al. (2011) aplicaram PLS e (LDA, do inglês: *linear discriminant analysis*) junto com algoritmo de projeções sucessivas (SPA, do inglês: *successive projections algorithm*) em dados de NIR para classificação de amostras de diesel em quatro classes, uma delas contendo amostras de diesel com 5% (v/v) de biodiesel. Khanmohammadi et al. (2013) utilizaram algoritmos genéticos (GA, do inglês: *genetic algorithm*) e SPA para processamento de dados espectrais visando a classificação de gasolina em amostra conforme ou não-conforme de acordo com 9 parâmetros de qualidade.

Em algumas aplicações não é necessário o conhecimento do valor exato da propriedade, somente se o biodiesel está conforme ou não-conforme segundo uma determinada especificação para aquela propriedade. Para estes casos seria mais apropriado em vez de utilizar métodos de predição e em seguida classificar o biodiesel a partir do valor predito, fazer diretamente a classificação a partir dos dados de entrada, minimizando o erro de predição.

Apesar de que métodos de regressão podem ser utilizados para prever parâmetros de qualidade de biodiesel e, então, classificá-lo quanto à sua conformidade, são escassos na literatura os trabalhos que aplicam diretamente métodos de classificação para essa finalidade. O presente trabalho otimizou e comparou quatro métodos (SVM, KNN, (TREE, do inglês: *Decision Tree Classifier*) e ANN) para a classificação direta de biodiesel de acordo com sua conformidade em relação a viscosidade, massa específica, estabilidade oxidativa e índice de iodo, a partir do seu perfil de FAME. As classificações foram baseadas nas especificações ASTM D6571, EN 14214 e RANP 45/2014.

1.2 Objetivos

1.2.1 Objetivo geral

Classificar amostras de biodiesel quanto a conformidade de acordo com as especificações vigentes em cada região (ASTM D6571, EN 14214 e RANP 45/2014) a partir dos percentuais de massas dos ésteres.

1.2.2 Objetivos específicos

- Obter uma base de dados de amostras de biodiesel;
- Encontrar os melhores classificadores para determinada propriedade e especificação;
- Fornecer uma ferramenta rápida e barata para que pequenos e médios produtores possam adequar sua produção.

1.3 Contribuições

O trabalho apresenta as seguintes contribuições:

- Avalia diferentes classificadores para a verificação da conformidade de biodiesel de acordo com as especificações ASTM D6571, EN 14214 e RANP 45/2014;
- Apresenta uma metodologia rápida e eficiente para a classificação de biodiesel quanto a conformidade de acordo com uma dada norma, levando então à economia de tempo, recursos financeiro e humano;
- Apresenta uma alternativa para a redução dos custos com mistura de diferentes matérias-primas, mantendo a conformidade do biodiesel.

1.4 Organização do Trabalho

No Capítulo 2, são apresentados os conceitos para o entendimento do problema. Em seguida é apresentada a base de dados e como foram feitos os experimentos para a verificação dos classificadores no Capítulo 3. No Capítulo 4 são apresentados os resultados obtidos e feita a discussão. Finalmente, no Capítulo 5 são feitas as conclusões, feita

uma análise das vantagens e desvantagens do método proposto e mostrados os trabalhos futuros.

Capítulo 2

Fundamentação Teórica

2.1 Biodiesel

O biodiesel compreende um composto de diesel e óleo vegetal ou gordura animal consistindo de longas cadeias de alquilésteres (metila, propila ou etila) (AHMAD *et al.*, 2012). Pode ser produzido a partir de uma grande variedade de matérias-primas das quais incluem óleos vegetais comuns como soja, algodão, dendê, canola, girassol e coco, como também gorduras animais (geralmente o sebo é utilizado) e ainda óleos de rejeito como por exemplo o óleo de fritura. A escolha da matéria-prima depende principalmente da geografia. Dependendo da origem e qualidade da matéria-prima algumas mudanças no processo de produção (KNOTHE; STEIDLEY, 2005; AHMAD *et al.*, 2012).

O biodiesel é miscível com o diesel em qualquer proporção (KNOTHE; STEIDLEY, 2005). As blends com diesel e biodiesel são denotados pelo acrônimo BX, onde a letra B indica que é uma blenda e X corresponde um número inteiro que representa a porcentagem de biodiesel na mistura.

É tecnicamente competitivo com outros combustíveis, oferecendo diversas vantagens sobre o diesel. Os óleos vegetais ou de origem animal são convertidos em ésteres (metílicos ou etílicos, dependendo do álcool utilizado) pelo processo de transesterificação na presença de um catalizador. O objetivo da transesterificação é diminuir a viscosidade do óleo, sendo o modo mais barato de transformar grandes estruturas moleculares ramificadas em moléculas menores de cadeia linear de modo a ser utilizada em motores de combustão diesel sem modificações (DEMIRBAS, 2008).

A transesterificação é a reação de substituição do álcool de um éster por outro em um processo similar à hidrólise, exceto que o álcool é usado no lugar da água. O álcool mais utilizado neste processo é o metanol, devido ao baixo custo e baixo teor de água. O

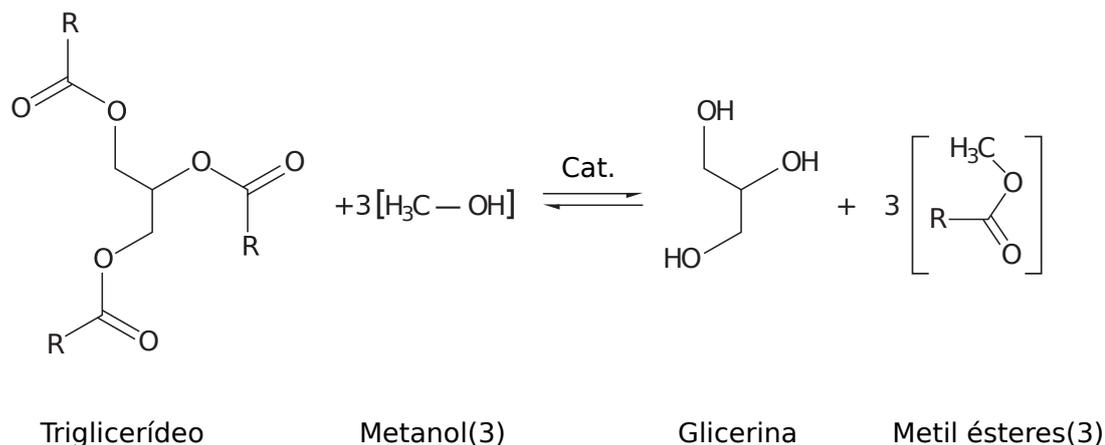


Figura 2.1: Reação de transesterificação de triglicerídeos. Adaptada de *Ahmad et al. (2012)*

processo é amplamente aplicado na produção de biodiesel por reduzir a alta viscosidade dos triglicerídeos (AHMAD et al., 2012). Outro álcool que pode ser utilizado é o etanol, com produção consolidada no Brasil, sendo menos tóxico, renovável e produz um biodiesel com maior número de cetano e lubrificidade (LÔBO; FERREIRA; CRUZ, 2009). A reação de transesterificação pode ser observada na Figura 2.1.

O biodiesel é caracterizado pelas suas propriedades físicas e de combustível, que incluem: massa específica, viscosidade, índice de iodo, ponto de nuvem e volatilidade. Produz menos potência e torque e também maior consumo que o diesel comum. Porém é melhor que o diesel em termos de teor de enxofre, ponto de fulgor, conteúdo aromático e biodegradabilidade (BALA, 2005; DEMIRBAS, 2008).

De acordo com a ANP, em abril de 2017 a principal matéria-prima utilizada para a produção de biodiesel no Brasil foi a soja, compondo 75,02% (Figura 2.2) (ANP, 2017). Algumas razões para a preferência pela soja no Brasil são: possui uma cadeia produtiva bem estruturada, alta tecnologia utilizada na produção, vasta rede de pesquisa, cultivo eficiente em todo o território, retorno rápido de investimento (4 a 5 meses), facilidade de comercialização, baixo custo (mais caro somente que algodão e gordura animal) entre outras (DALL'AGNOL, 2015). Mesmo assim, a produtividade obtida com o uso da soja não é muito favorável, produzindo algo em torno de 700 l/ha, cerca de 1/3 do que é produzido com outras matérias-primas (COSTA; JUNIOR; ARANDA, 2010).

2.1.1 Benefícios do biodiesel e conformidade com os padrões industriais

Por ser uma energia renovável, que substitui os derivados do petróleo e podendo ser utilizado em misturas de até 20% sem qualquer alteração nos motores convencionais,

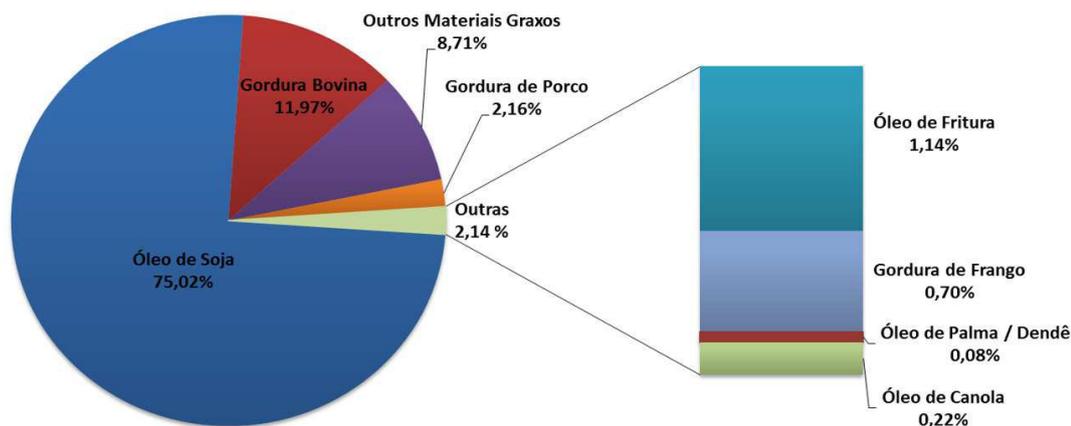


Figura 2.2: *Matérias-primas utilizadas para produção de biodiesel no Brasil (Abril/2017). Adaptada de ANP (2017)*

pode reduzir as emissões de escapamentos e não é tóxico, biodegradável e aceito para ambientes sensíveis. Estes benefícios relacionados à indústria de padrões segue (DAHIYA, 2014; NREL, 2008):

Fácil de usar: Misturas de 20% ou menos não requerem nenhum equipamento adicional ou qualquer modificação pois o biodiesel B20 pode ser armazenado em tanques de diesel e bombeados como o mesmo equipamento.

Melhora a operação do motor: uma melhora na lubrificação (necessária para evitar o desgastes de partes móveis como bombas de combustível) e aumento do número de cetano do combustível mesmo em baixas concentrações. A redução de emissões de enxofre para 15 ppm pelas regulamentações também reduz a lubrificação do diesel e os processos de hidrotreamento usados para reduzir o componentes aromáticos e enxofre também reduz impurezas polares como componentes nitrogenados, que melhoram a lubrificação.

Efeitos positivos do biodiesel nos gases tóxicos e melhoria da saúde humana: Algumas partículas e emissões provenientes do diesel são tóxicas e até cancerígenas. O uso de biodiesel B100 pode eliminar em 90% esses gases tóxicos e a utilização de B20 gera uma redução de 20-40%.

2.1.2 Propriedades

As propriedades físico-químicas mais importantes que influenciam o desempenho de motores, meio ambiente e características econômicas são tempo de estabilidade química e térmica, densidade, viscosidade, velocidade do som, modulus bulk, número de cetanos, ponto de congelamento, ponto de fluidez, ponto de entupimento de filtro a frio, ponto de fulgor, tendência de entupimento de filtro, corrosividade, lubrificação, composi-

ção química, teor de contaminantes e teor de água (KEGL; KEGL; PEHAN, 2013). Neste trabalho foram testadas quatro propriedades viscosidade, índice de iodo, estabilidade oxidativa e densidade que estão relacionadas com a composição e por isso podem ser inferidas a partir dos percentuais de massas dos ésteres.

A viscosidade mede a resistência de um líquido em fluir devido à fricção interna causada por uma parte do fluido mover sobre a outra. Esta propriedade é uma propriedade crítica, por afetar o comportamento da injeção de combustível. Viscosidade muito alta leva a atomização de combustível pobre e pode causar um maior tamanho das gotas, pobre vaporização, ângulo do spray de injeção mais estreito e maior penetração do spray de combustível no cilindro. Uma baixa viscosidade pode causar um excessivo desgaste nas bombas de injeção e perda de potência devido a vazamento da bomba (SARIN, 2012).

Índice de iodo é um índice do número de ligações duplas no biodiesel, com isso é o parâmetro que quantifica o grau de insaturação do biodiesel. É expresso como a quantidade de iodo em gramas que reage com 100 g de biodiesel. Foi originalmente inserida na especificação europeia EN14214 para garantir uma estabilidade oxidativa satisfatória, contudo é uma medida simples do total de insaturação enquanto que a estabilidade oxidativa é mais fortemente influenciada pela quantidade de moléculas FAME tendo múltiplas ligações duplas.

A estabilidade oxidativa do biodiesel depende da sua composição de ácidos graxos bem como de antioxidantes naturais. Biodiesel com alto teor de ácidos graxos insaturados que são mais sensíveis à degradação oxidativa levando a baixos valores de período de indução. O período de indução mede o quanto o biodiesel é suscetível à degradação tanto térmica quanto oxidativa, formando produtos insolúveis que ocasionam problemas de formação de depósitos e entupimento do sistema de injeção de combustível.

A massa específica do biodiesel está relacionada à estrutura molecular. É diretamente proporcional ao comprimento da cadeia carbônica do alquiléster e inversamente proporcional ao número de insaturações. É uma propriedade que afeta o desempenho do motor. Como a bomba de injeção de combustível mede o combustível pelo volume e não pela massa, uma maior ou menor quantidade de massa de combustível é injetada dependendo da sua densidade (HOEKMAN *et al.*, 2012).

A determinação destas propriedades por métodos experimentais é um processo caro e demorado. Com isso a obtenção de um modelo que permita com precisão a determinação dessas propriedades é de grande ajuda para a comunidade científica. Alguns trabalhos apresentados na literatura fizeram a predição de propriedades do biodiesel utilizando modelos empíricos (CHAVARRIA-HERNANDEZ; PACHECO-CATALÁN, 2014), MLR (RAMÍREZ-VERDUZCO; RODRÍGUEZ-RODRÍGUEZ; JARAMILLO-JACOB, 2012; SA-

XENA; JAWALE; JOSHIPURA, 2013; HONG; JEON; LEE, 2014) e ANN (YANG et al., 2002; MENG; JIA; WANG, 2014).

Em algumas aplicações não é necessário o conhecimento do valor exato da propriedade, somente se o biodiesel está conforme ou não-conforme segundo uma determinada especificação para aquela propriedade. Para estes casos é mais apropriado em vez de utilizar métodos de predição e em seguida classificar o biodiesel a partir do valor predito, fazer diretamente a classificação a partir dos dados de entrada, minimizando o erro de predição.

2.1.3 Padrões para Caracterização do Biodiesel

O controle de qualidade durante o processo de produção e distribuição do biodiesel é um fator crítico para pesquisadores e usuários finais que precisam de uma forma de avaliação da qualidade consistente que garante um bom desempenho dos motores. A qualidade e eficiência do biodiesel pode ser medida através de algumas propriedades e é regulamentada por especificações adotadas em alguns países.

As propriedades do biodiesel vão depender da composição dos ácidos graxos do óleo utilizado como matéria-prima. As variações nas proporções destes ácidos graxos é a razão pela qual as normas fornecem intervalos para algumas propriedades como a densidade.

O primeiro país a determinar um padrão de qualidade a ser adotado para biodiesel foi a Áustria com a norma ON C1190 de 1990 padronizando o éster metílico de óleo de colza. Os padrões indicam quais propriedades devem ser medidas, as faixas que elas devem estar para atender as especificações, unidades de medidas e normas que especificam como as propriedades devem ser medidas. Os padrões podem indicar valores máximos (por exemplo para impurezas como água e sedimentos) e/ou valores mínimos (período de indução, viscosidade) (ROMANO; SORICHETTI, 2010).

Os padrões atuais de qualidade do biodiesel são baseados em fatores que variam de região para região, como: características dos padrões de diesel existentes, os tipos de motores mais comuns, controles governamentais de regulação de emissões, grau de desenvolvimento e fatores climáticos e/ou o propósito e motivação da utilização do biodiesel. Na Europa a quantidade de carros que utilizam motor diesel é considerável, enquanto que nos Estados Unidos e no Brasil o diesel é utilizado principalmente em caminhões e veículos pesados. as matérias-primas mais utilizadas na Europa são colza e girassol, nos Estados Unidos e Canadá soja e óleo de fritura, enquanto que no Brasil a principal é a soja, portanto é de se esperar que sejam encontradas diferenças significativas entre os padrões adotados em cada região (BARABÁS; TODORUȚ, 2011; SARIN, 2012).

A qualidade do biodiesel pode variar conforme os ésteres disponíveis no óleo utilizado como matéria-prima, de acordo com o processo de produção ou mesmo durante a estocagem. As estruturas moleculares dos ésteres podem mudar de tamanho da cadeia carbônica e também na quantidade e posição de insaturações ou agrupamentos na cadeia (LÔBO; FERREIRA; CRUZ, 2009).

Os parâmetros utilizados para caracterizar a qualidade do biodiesel podem ser classificados em dois grandes grupos: o primeiro incluindo propriedades gerais como massa específica, viscosidade, ponto de fulgor, ponto de fluidez e o segundo grupo descreve a composição química e pureza da mistura de ésteres de ácidos graxos (teor de álcool, teor de éster, monoacilglicerol, diacilglicerol, triacilglicerol) (MEHER; SAGAR; NAIK, 2006; ROMANO; SORICHETTI, 2010).

2.1.3.1 EN 14214

O padrão EN 14214 é adotado por todos os 31 países membros do Comitê Europeu de Normalização (CEN): Áustria, Bélgica, Bulgária, Croácia, Chipre, República Checa, Dinamarca, Estônia, Finlândia, França, Alemanha, Grécia, Hungria, Islândia, Irlanda, Itália, Letônia, Lituânia, Luxemburgo, Malta, Holanda, Noruega, Polônia, Portugal, Romênia, Eslováquia, Eslovênia, Espanha, Suécia, Suíça e Reino Unido. Dessa forma não há regulamentações nacionais sobre a qualidade do biodiesel, mas quanto ao ponto de congelamento. As organizações de padronização nacionais fornecem as especificações para regulações do ponto de entupimento (CCPP, do inglês *cold-filter plugging point*), viscosidade, massa específica e destilação que dependem do clima (BARABÁS; TODORUȚ, 2011).

2.1.3.2 ASTM D6751

O padrão de qualidade de biodiesel americano é controlado pela Sociedade Americana para Ensaio de Materiais (ASTM, do inglês *American Society of Testing and Materials*) através da norma ASTM D6751. Qualquer biodiesel usado nos Estados Unidos para mistura deve estar primeiramente conforme com a norma ASTM D6751, aplicada para ésteres metílicos e etílicos (BARABÁS; TODORUȚ, 2011; NREL, 2008).

Como qualquer outro padrão de combustíveis da ASTM, a ASTM D6751 é baseada em propriedades químicas e físicas necessárias para a operação segura e satisfatória de um motor diesel. Não é baseada em uma matéria-prima específica ou algum processo de produção de biodiesel. O produto final deve estar com as propriedades especificadas conforme a norma e de acordo com a seguinte definição: biodiesel, combustível constituído

de mono-alkilésteres de ácidos graxos de cadeia longa, derivados de óleos vegetais ou gorduras animais, designado B100 (NREL, 2008).

As especificações contidas não foram destinadas a ser aplicadas a um biodiesel B100 na forma pura. Contudo, os compradores e vendedores são incentivados a utilizar a norma ASTM D6751 para comercialização de biodiesel na forma B100 ou na forma de mistura. Outros arranjos ou especificações podem ser legalmente utilizados caso as duas partes da compra concordem, obedecendo as regulamentações federais e estaduais (limites de enxofre da agência de proteção ambiental, a agência de administração e segurança no trabalho limita o ponto de fulgor, etc.) (NREL, 2008).

2.1.3.3 Resolução ANP 45/2014

No Brasil as especificações do biodiesel são regulamentadas pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), através da Resolução ANP 45/2014 (RANP 45/2014) (ANP, 2014). Os padrões estabelecidos utilizaram como base as normas europeia EN 14214 e americana ASTM D6751.

Na Tabela 2.1 pode ser vista uma comparação entre as especificações EN 14214, ASTM D6751 e RANP 45/2014. Pode ser observado que EN 14214 é a que requer a determinação de mais parâmetros que a ASTM D6751, e a RANP 45/2014 estando no meio termo.

A RANP 45/2014 é mais exigente quanto ao teor de água (200 mg/kg contra 500 mg/kg nas outras especificações, sendo portanto menos da metade). A necessidade dessa restrição é devido às particularidades da logística no território nacional. Atender a esse parâmetro exige investimentos nas fábricas brasileiras para obtenção de um biodiesel com baixo teor e também para manter durante transporte e armazenamento (diferentes critérios tolerância são adotados para fiscalização de produtor e distribuidor) (BRASIL, 2015).

		Brasil	Estados Unidos	União Europeia	Canadá
Instrumento Normativo		Resolução ANP 45/2014	ASTM D6751-15	EN14214:2012	CAN/CGSB-3524-2011
Denominação		B100	Grade 1-B e 2-Bd	FAME	B100
Ano de implementação		2014	2015	2014	2011g
Parâmetros	Unidade				
Aspecto	-	LIIa	-	-	-
Massa específica a 15°C, min.	-	máx.	kg/m ³	850-900	(20°C)
Viscosidade cinemática a 40°C, min.	-	máx.	mm ² /s	3,0-6,0	1,9-6,0
Água e sedimentos, máx.	%vol	-	0,05	-	-
Teor de água, máx.	mg/kg	200b	-	500	400
Contaminação total,	máx.	mg/kg	24	-	24
Ponto de fulgor	°C	100	93	101	130
Temperatura Destilação	90%	recuperado,	máx	°C	-
Teor de éster	%massa	96,5	-	96,5	-
Teor de éster linolênico	%massa	-	-	12,0	-
Teor de ésteres poliinsaturados	%massa	-	-	1,0	-
Resíduo de carbono,	máx.	%massa	-	0,05	-
Cinzas sulfatadas,	máx.	%massa	0,02	0,02	0,02
Enxofre total,	máx.	mg/kg	10	15	/
Teor de Sódio	+	Potássio,	máx.	mg/kg	5
Teor de Cálcio	+	Magnésio,	máx.	mg/kg	5
Teor de fósforo,	máx.	mg/kg	10	10	4,0
Corrosividade ao cobre, 3h a 50°C	Grau	1	3	1	-
Número de cetano	-	Anotar	47	51	Anotar
Ponto de entupimento	a	frio,	máx.	°C	c
Ponto de névoa	°C	-	Anotar	-	Anotar
Filtrabilidade a frio	s	-	d	-	-
Índice de acidez,	máx.	mg	KOH/g	0,5	
Glicerina livre,	máx.	%massa	0,02	0,02	0,02
Glicerina total,	máx.	%massa	0,25	0,24	0,25
Monoglicerídeos,	máx.	%massa	0,7	0,4	-d
Diglicerídeos,	máx.	%massa	0,2	-	0,2
Triglicerídeos,	máx.	%massa	0,2	-	0,2
Metanol / Etanol, máx.	%massa	0,2	0,2e	0,2	0,20
Índice de iodo, máx	g	de	I ₂ /100g	Anotar	-
Estabilidade à oxidação a 110°C, min.	Horas	8	3	8	8

Tabela 2.1: Comparação entre as propriedades especificadas pelas normas RANP 45/2014, ASTM D6751, EN 14214 e CAN/CGSB-3524-2011. Adaptada de *Brasil (2015)*

a. O B100 somente poderá ser considerado não conforme após avaliação do teor de água e contaminação total.

b. Para fins de Fiscalização são adotados os seguintes valores: 250 e 350 mg/kg para o produtor e distribuidor, respectivamente.

c. Por possuir grande variação regional são adotados limites diferenciados por estado ao longo do ano (Tabela II da RANP 45/2014).

d. A ASTM D6571 estabelece que o Grade 1B deve apresentar filtrabilidade a frio (ASTM D7501) de 200 e teor máximo de monoglicerídeos (ASTM D6584) de 0,4%. Já o Grade 2B deve apresentar 360 para o ensaio de filtrabilidade a frio e sem limite máximo para monoglicerídeos.

e. Se a amostra apresentar ponto de fulgor superior a 130 °C não é necessário realizar o ensaio de teor de metanol/etanol. f. Tendo em vista a utilização do FAME em uma ampla faixa de temperaturas dentro da União Europeia, a CEN estabeleceu seis especificações de fluidez a frio (Grade A - F) com valores de ponto de entupimento a frio variando de +5 a -20, respectivamente. Para utilizações em climas árticos são definidos adicionalmente as classes 0 - 4 que devem apresentar valores variando de -20 a -44°C, respectivamente.

g. A norma que estabelece a especificação do B100 no Canadá (CGSB CAN/CGSB-3.524) sofreu uma revisão em 2014. No entanto, a ANP não possui acesso a esta versão da norma, por isso foram reportados dados referentes a versão de 2011

A EN 14214 apresenta alguns parâmetros específicos (ácido linolênico, metil esterres com mais de 4 insaturações), que estão mais relacionados quanto à matéria-prima que ao processo de produção, relacionadas com a estabilidade oxidativa (já este parâmetro está contido na RANP 45/2014 e com o mesmo valor que a norma europeia). Como as matérias-primas usadas no Brasil não possuem altas concentrações de ácidos linolênicos estes parâmetros não são restritivos para o biodiesel brasileiro (BRASIL, 2015).

Quando companhias de óleo diesel compram biodiesel para misturar com diesel eles exigem especificações restritas. Muitas companhias da União Europeia exigem os seguintes valores (ROMANO; SORICHETTI, 2010):

- Teor de água < 300 ppm
- Contaminação total < 15 ppm
- Monoglicerídeo < 0,4% w/w
- Estabilidade oxidativa > 8 h
- FBT (tendência de entupimento de filtro pela norma IP387) <2
- Esterol-glicosídeos < 20 ppm.

2.1.4 Controle de qualidade em diferentes escalas de produção

O caso ideal seria que todas as propriedades estabelecidas na norma vigente pudessem ser medidas para verificar se elas atendem as exigências. Esta é a única forma de garantir que o produto final seja realmente biodiesel. Este processo é aplicado quando o biodiesel é para exportação ou para consumo doméstico. Em plantas que produzem grandes volumes de biodiesel, o controle de qualidade é contínuo e as amostras são analisadas em um laboratório próprio ou periodicamente enviadas para laboratórios contratados (ROMANO; SORICHETTI, 2010).

Por outro lado, pequenos produtores não podem dispor de despesas e pessoal para esse tipo de controle. Com volumes de produção mais baixos, como para consumo próprio ou pequenas e médias companhias o controle de qualidade é feito em lotes e então a qualidade pode variar dependendo do lote. Um laboratório próprio é inviável e o envio periódico de amostras é também bastante custoso. Mesmo assim estes pequenos produtores tendem a assumir que o seu produto final é biodiesel, o que é uma suposição incorreta para a maioria dos casos. O problema da falta de controle de qualidade em pequenos e médios produtores é um sério problema global, não se restringindo a alguns países (ROMANO; SORICHETTI, 2010).

2.1.5 Viabilidade Econômica do Biodiesel

A viabilidade econômica da utilização do biodiesel depende logicamente do preço do petróleo cru e dos custos de transporte do diesel para locais de consumo distantes. O custo do petróleo tende a aumentar devido ao aumento da demanda e oferta limitada. Além disso limitações quanto ao teor de aromáticos e enxofre irão resultar em altos custos na produção de diesel (AHMAD et al., 2012).

O custo da produção de ésteres metílicos a partir de óleos comestíveis é atualmente mais caro que o diesel comum. Devido aos custos relativamente altos dos óleos vegetais, os ésteres metílicos produzidos não podem competir diretamente com o diesel comum sem a ajuda de subsídios do governo (DEMIRBAS, 2008; AHMAD et al., 2012).

Atualmente, a maior parte do biodiesel produzido no mundo tem como matéria-prima óleos comestíveis disponíveis em larga escala pela agroindústria (cerca de 95%) (GUI; LEE; BHATIA, 2008).

Os custos do biodiesel podem reduzir consideravelmente com a utilização de óleos não comestíveis como óleo de fritura, sebo, gordura de porco, mahua, karanja, pinhão-manso, mamona e outros (DEMIRBAS, 2005; AHMAD et al., 2012; ATABANI et al., 2013; YUSUP; KHAN, 2010). Segundo Gui, Lee e Bhatia (2008) o óleo de fritura de soja era vendido por aproximadamente US\$1,09/galão no ano de 2013, enquanto que o óleo de soja era vendido por US\$2,22/galão. Uma grande diferença em relação ao diesel mineral é a baixa emissão de gases do efeito estufa, com uma diminuição de 50% na geração de monóxido de carbono e 60% de fuligem, menor toxicidade e isenção de enxofre (BARNWAL; SHARMA, 2005; CREMONEZ et al., 2015).

Em relação aos óleos vegetais, as gorduras animais como sebo, gordura de frango e gordura amarela são frequentemente mais baratas para a produção de biodiesel. O problema da limitação do uso das gorduras animais é que elas nunca poderão suprir as necessidades mundiais. Uma outra desvantagem é que as gorduras animais criam um grande problema na produção de biodiesel: tornam-se sólidas na temperatura ambiente (WEN; JOHNSON, 2009; ROJAS-GONZÁLEZ; GIRÓN-GALLEGO, 2011; BANKOVIĆ-ILIĆ; STAMENKOVIĆ; VELJKOVIĆ, 2012). Isso dificulta a sua aplicação em climas frios por conta do seu alto teor de ácidos graxos saturados (CREMONEZ et al., 2015).

Como observado por (GUI; LEE; BHATIA, 2008), a competição da utilização do óleo de dendê mundialmente utilizado como óleo comestível com a produção de biodiesel causou o aumento do preço do óleo cru significativamente. Uma clara tendência de que os óleos comestíveis não são as melhores alternativas para a produção de combustível.

Dal Belo Leite et al. (2015) mostraram que mamona e girassol são economicamente

competitivos com milho na região de Montes Claros, porém, necessidades alimentares e de trabalho dificultem a adoção dos agricultores. O problema da utilização de óleos de baixo custo e gorduras é que eles apresentam grandes quantidades de ácidos graxos livres e não podem ser convertidos em biodiesel utilizando um catalizador alcalino.

Segundo [Costa et al. \(2013\)](#) a utilização de resíduos seria a melhor alternativa em termos sócio-econômicos e ambientais, não podendo ser adotada por diferentes razões. A soja não é o óleo mais barato, ou que gera mais empregos e também agrega menos benefícios comparada a resíduos e corresponde pela maior parte da produção brasileira devido a sua cadeia produtiva consolidada.

É observado (Figuras 2.3 e 2.4) que a demanda por biodiesel vem crescendo com os anos e que a produção nacional está ligada a essa demanda. As regiões nordeste, norte e sul têm as menores produções e não atendem ao consumo, enquanto que as regiões centro-oeste e sul consomem bem menos do que produzem e ainda contam com uma capacidade nominal maior que duas vezes a produção atual.

O Brasil lidera a produção de biodiesel na América Latina, tendo começado nos anos 70, não por questões ambientais, mas na busca pela auto-suficiência energética. Na década seguinte o governo brasileiro implementou o Programa de Óleos Vegetais (OVEG), onde foi estudada a utilização de misturas com óleo mineral em diferentes proporções ([CREMONEZ et al., 2015](#)).

Ao contrário de outros países, o programa de biodiesel brasileiro é utilizado pelo governo como forma de inclusão da agricultura familiar, utilizando culturas adaptadas a condições específicas de cada região do país, permitindo a inclusão de matérias-primas de cultivo dos produtores locais.

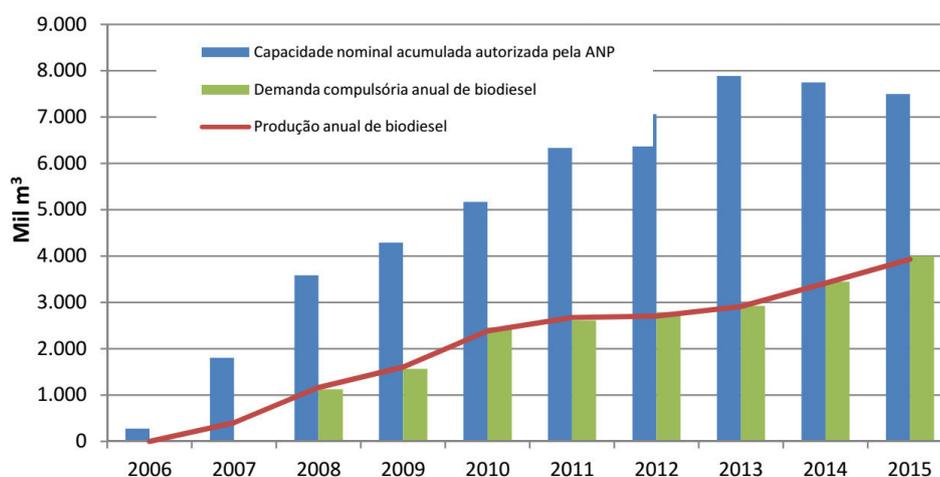


Figura 2.3: Evolução anual da produção, demanda compulsória e capacidade nominal. Adaptada de ANP (2017)

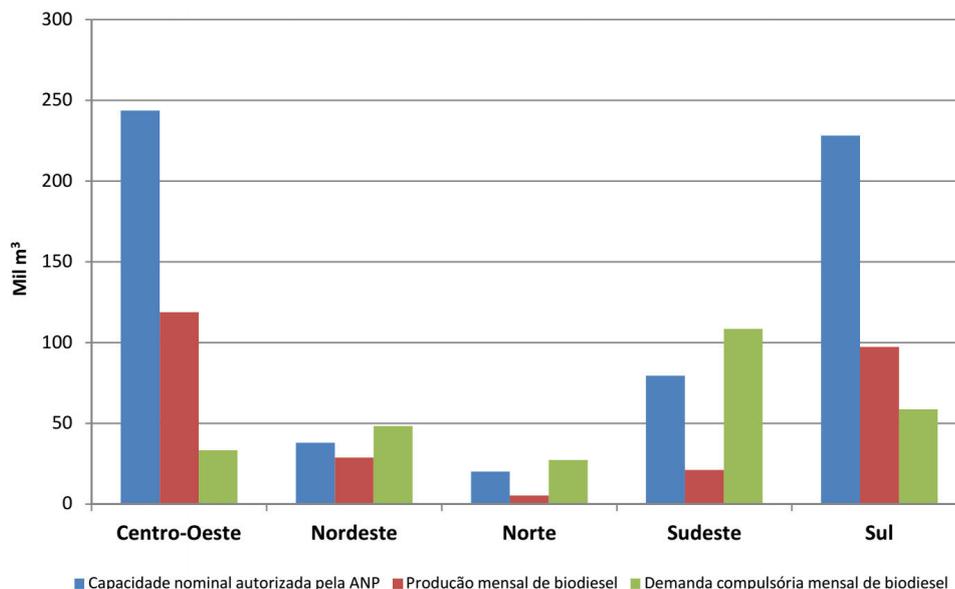


Figura 2.4: *Produção, demanda compulsória e capacidade nominal por região. Adaptada de ANP (2017)*

2.2 Classificadores

A classificação é a tarefa de aprender uma função objetivo f que mapeia cada atributo de um conjunto \mathbf{x} para um dos rótulos de classe predefinido y . A função objetivo também pode ser chamada de modelo classificador (TAN; STEINBACH; KUMAR, 2006).

Neste trabalho foram testados 3 algoritmos para classificação que são avaliados de acordo com seu desempenho para cada propriedade e especificação.

2.2.1 Máquina de Vetor de Suporte

A máquina de vetor suporte (SVM, do inglês *support vector machine*) foi introduzida por Vapnik (1995), podendo ser utilizada tanto para problemas de classificação (ALVES; POPPI, 2013; SHI et al., 2013; BRUDZEWSKI et al., 2006) quanto de predição (ALVES; HENRIQUES; POPPI, 2012; FILGUEIRAS et al., 2014). É um método de aprendizagem supervisionado capaz de identificar a partir de n indivíduos pertencentes a diversos subgrupos a que classe um dado indivíduo deve pertencer (BISHOP, 2006).

Para efeito de simplificação, considerando o caso de duas classes usando um modelo linear da seguinte forma:

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0 \quad (2.1)$$

sendo \mathbf{w} um vetor de pesos ajustados, b é o bias e \mathbf{x} é o vetor de treinamento de entrada, com seus respectivos rótulos $y_i \in Y$, tendo $Y = \{-1, 1\}$. O modelo definido na Equação 2.1 define um hiperplano ótimo que classifica todos os dados de entrada, caso \mathbf{x} seja linearmente separável todos os dados das classes -1 e 1 serão divididos por esse hiperplano (SCHÖLKOPF; SMOLA, 2002). Para ajudar a ilustrar o hiperplano ótimo para uma entrada linearmente separável pode ser observada na Figura 2.5.

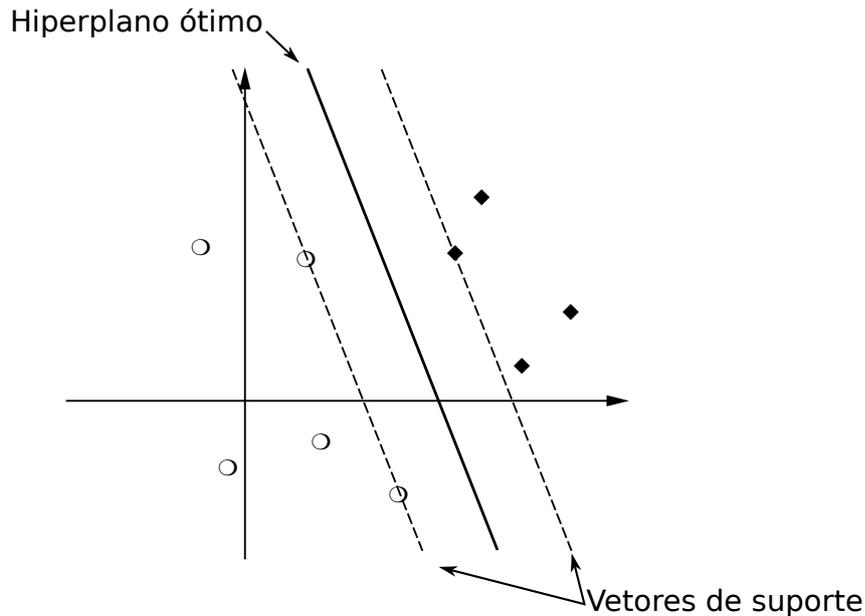


Figura 2.5: Conjunto de treinamento contendo duas classes com vetores de suporte e hiperplano ótimo. Adaptado de Schölkopf e Smola (2002)

$$\begin{aligned} y(\mathbf{x}) = 1 &\rightarrow \mathbf{w}^T \mathbf{x} + b \geq 0, \\ y(\mathbf{x}) = -1 &\rightarrow \mathbf{w}^T \mathbf{x} + b < 0. \end{aligned} \quad (2.2)$$

Como exemplificado na Figura 2.6, sendo x_1 um ponto pertencente ao hiperplano $H_1 : \mathbf{w}^T \mathbf{x} + b = 1$ e x_2 um ponto pertencente ao hiperplano $H_2 : \mathbf{w}^T \mathbf{x} + b = -1$.

A projeção de $x_1 - x_2$ na direção de \mathbf{w} , perpendicular ao hiperplano ótimo permite obter a distância entre os hiperplanos H_1 e H_2 , dada por

$$(x_1 - x_2) \left(\frac{\mathbf{w}}{\|\mathbf{w}\|} \frac{(x_1 - x_2)}{\|x_1 - x_2\|} \right) \quad (2.3)$$

De acordo com as Equações 2.2, tem-se que $\mathbf{w}^T x_1 + b = 1$ e $\mathbf{w}^T x_2 + b = -1$. A diferença entre as equações dá $\mathbf{w}(x_1 - x_2) = 2$ e substituindo o resultado em 2.3:

$$\frac{2}{\|\mathbf{w}\|} \frac{(x_1 - x_2)}{\|x_1 - x_2\|} \quad (2.4)$$

e tomando-se a norma, obtém-se

$$\frac{2}{\|\mathbf{w}\|} \quad (2.5)$$

Esta é a distância d , ilustrada na Figura 2.6, que separa os hiperplanos H_1 e H_2 , paralelos ao hiperplano ótimo. A minimização de $\|\mathbf{w}\|$ permite maximizar a margem de separação das classes em relação a $\mathbf{w}^T \mathbf{x} + b = 0$. Dessa forma tem-se o problema de otimização:

$$\min_{w,b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.6)$$

restringido a: $y_i(\mathbf{w}^T x_i + b) - 1 \geq 0, \quad \forall i = 1, 2, \dots, n$

Estas restrições asseguram que não haja dados de treinamento entre as margens de separação das classes. Trata-se de um problema de otimização quadrático, com função objetivo convexa cujos pontos que satisfazem as restrições formam um conjunto convexo, possuindo portanto um único mínimo global.

Para a obtenção da solução aplica-se um operador Lagrangiano, que engloba as restrições às funções objetivo, associadas aos multiplicadores de Lagrange, de acordo com:

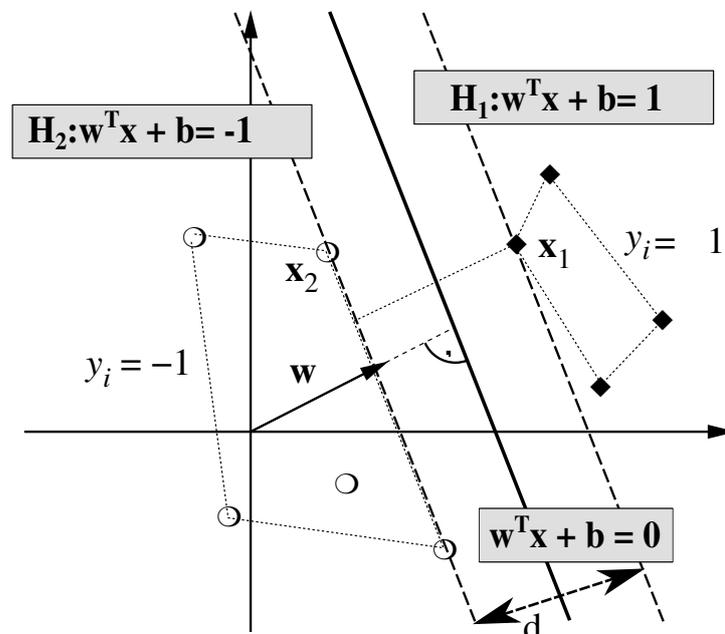


Figura 2.6: Conjunto de treinamento contendo duas classes com vetores de suporte e hiperplano ótimo. Adaptado de *Schölkopf e Smola (2002)*

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w}^T x_i + b) - 1) \quad (2.7)$$

onde a minimização do operador Lagrangiano implica na maximização das variáveis α_i e na minimização de w e b .

Fazendo as derivadas de L em relação a b e w iguais zero,

$$\frac{\partial}{\partial b} L(\mathbf{w}, b, \alpha) = 0 \text{ e } \frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, b, \alpha) = 0 \quad (2.8)$$

obtem-se

$$\sum_{i=1}^m \alpha_i y_i = 0 \quad (2.9)$$

e

$$\mathbf{w} = \sum_{i=1}^m \alpha_i y_i x_i \quad (2.10)$$

Substituindo as Equações 2.9 e 2.11 na Equação 2.7 tem-se o seguinte problema de otimização:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (2.11)$$

sujeito a $\alpha_i \geq 0 \forall i = 1, \dots, m$ e $\sum_{i=1}^m \alpha_i y_i = 0$ sendo que $(x_i \cdot x_j)$ corresponde ao produto interno de x_i e x_j

Quando se trata de dados não linearmente separáveis, não é possível a construção de um hiperplano ótimo que separe as classes sem erros de classificação (HAYKIN; LIPPMANN, 1999). Para estes casos, é introduzido um conjunto de variáveis escalares não negativas, ξ_i , na definição do hiperplano de separação:

$$d_i(\mathbf{w}^T x_i + b) \geq 1 - \xi_i, \text{ para } i = 1, 2, \dots, n \quad (2.12)$$

As variáveis ξ_i são chamadas de variáveis de folga, medem o desvio de um ponto dado da condição ideal de separabilidade de padrões. O problema passa a ser então:

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i \quad (2.13)$$

sujeito a: $y_i(\mathbf{w} \cdot x_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$ e $\forall i = 1, \dots, n$

Da mesma forma, trata-se de um problema de otimização quadrático com restrições lineares. Aplicando o operador Lagrangiano e tomando as derivadas parciais iguais a zero, tem-se o seguinte problema de otimização:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (2.14)$$

sujeito a: $0 \leq \alpha_i \leq C$, $\forall i = 1, \dots, n$ e $\sum_{i=1}^n \alpha_i y_i = 0$ onde C é um parâmetro positivo especificado pelo usuário.

Os problemas de classificação não lineares são resolvidos através do mapeamento do conjunto de treinamento para um novo espaço de maior dimensão, denominado espaço de características. Seja $\phi(x)$ uma função que mapeia o espaço de entrada para o espaço de característica, esta função é usada para mapear x_i e x_j antes da realização do produto interno entre eles:

$$k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad (2.15)$$

Esta função é chamada de função *kernel*. Modificando o problema de maximização para:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (2.16)$$

com as mesmas restrições da Equação 2.14.

Alguns exemplos de funções *kernel*

- Polinomial: $k(x, x') = (x \cdot x')^d$,
- Gaussiano: $k(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$
- Sigmoidal: $k(x, x') = \tanh(\kappa(x \cdot x') + \Theta)$

a escolha apropriadamente os parâmetros $d \in \mathbb{N}$ e $\sigma, \kappa, \Theta \in \mathbb{R}$ empiricamente levou a classificadores SVM muito similares em precisão, nesse sentido o conjunto de vetores de suporte parece caracterizar a tarefa de modo que de certa forma é independente do tipo de kernel usado, desde que os parâmetros do *kernel* sejam ajustados adequadamente (SCHÖLKOPF; SMOLA, 2002).

2.2.2 K-nearest Neighbor

O classificador K -NN foi descrito por Cover e Hart (1967). É um método de classificação não paramétrico muito popular devido a sua simplicidade e razoável eficiência. Seu desempenho depende fortemente da métrica usada para o cálculo da distância entre os pontos.

Supondo um conjunto D de amostras utilizadas no treinamento. Cada elemento de D é dado por $(x_1, x_2, \dots, x_n, c)$, sendo c a classe à qual o conjunto de características (x_1, x_2, \dots, x_n) pertence. Considerando uma nova amostra não classificada $Y = (y_1, y_2, \dots, y_n)$, para ser classificada calcula-se as distâncias de Y em relação a todas as amostras utilizadas no treinamento e considerando as k amostras de treinamento mais próximas a Y . A amostra Y será classificada na classe que ela está mais próxima.

A distância pode ser calculada por diversas métricas:

- Euclidiana: $d_{st}^2 = (x_s - y_t)(x_s - y_t)'$
- Mahalanobis: $d_{st}^2 = (x_s - y_t)C^{-1}(x_s - y_t)'$
- Minkowski: $d_{st} = \sqrt[p]{\sum_{j=1}^n |x_{sj} - y_{tj}|^p}$

De forma geral é preciso fazer a normalização das características de entrada de modo que todas apresentem o mesmo intervalo de variação, que poderia influenciar no cálculo da distância.

2.2.3 Árvore de Decisões

O funcionamento da árvore de decisão é baseado em uma série de questões sobre os atributos de entrada. Cada vez que uma resposta é fornecida, uma nova questão é feita até chegar à conclusão sobre a qual classe pertence o dado. A série com as questões e suas possíveis respostas pode ser organizada na forma de árvore de decisão, onde uma estrutura hierárquica formada por nós e ramos. A árvore de decisões tem três tipos de nós

- **nó raiz**, que não tem ramos de entrada e nenhum ou mais ramos de saída.
- **nós internos** cada nó que contém um ramo de entrada e dois ou mais ramos de saída.
- **nós terminais ou folhas** cada nós que tem um ramo de entrada e nenhum ramo de saída.

Para cada folha é atribuído um rótulo de classe. Os outros nós, que incluem a raiz e os nós internos contêm condições de teste de atributos para separar dados de diferentes características (diferentes classes). Tomando como exemplo a Figura 2.7, o atributo *Temperatura do corpo* é usado para dividir os vertebrados em sangue quente e sangue frio. Como todos os vertebrados de sangue frio não são mamíferos, então um nó folha é rotulado Não-mamífero é criado na ramificação direita do nó raiz. Se o vertebrado tem sangue quente, um atributo em seguida, *Dá à luz* é usado para diferenciar os mamíferos dos outros animais de sangue quente, como as aves.

A classificação de um novo dado se torna simples, quando a árvore de decisão já está construída. Começando do nó raiz, aplica-se as condições ao novo dado e segue pelo ramo apropriado baseado no resultado. Pode levar a um nó terminal com uma nova condição de teste ou a um nó folha. A classe associada ao nó folha é então associada ao novo dado.

Um algoritmo de aprendizagem de uma árvore de decisão trata duas questões:

1. **Como os dados de treinamentos devem ser particionados:** Em cada passo recursivo do algoritmo deve ser selecionado uma condição de teste para dividir os dados em subconjuntos menores. O algoritmo deve utilizar um método para especificar a condição de teste para diferentes tipos de atributos e ainda uma medida para avaliar a eficiência de cada condição de teste.
2. **Como deve ser o procedimento de parada do particionamento:** Uma condição de parada é necessária para terminar o processo de crescimento da árvore. Uma estratégia é continuar expandindo um nó até que todos os dados pertençam à

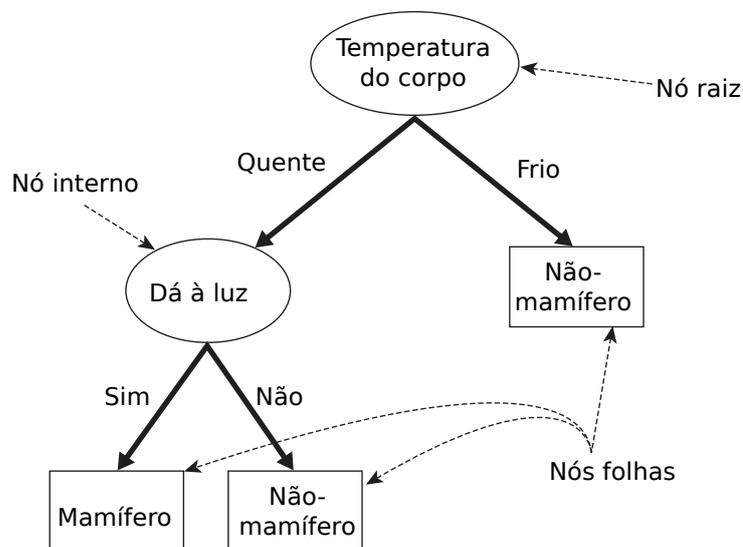


Figura 2.7: Árvore de decisão para um problema de classificação de mamífero. Adaptado de Tan, Steinbach e Kumar (2006)

mesma classe ou tenham valores de atributos idênticos. Embora estas duas condições sejam suficientes para a parada de uma árvore de decisão, outro critério pode ser implementado para reduzir o tempo necessário.

Diversas medidas são utilizadas para determinar o modo de partição dos dados em uma árvore de decisão e são definidas em termos da distribuição das classes nos dados antes e depois da partição.

Considerando $p(i)$ como a fração dos dados que pertencem à classe i , para um problema de duas classes a distribuição de classes em qualquer nó é dada por (p_0, p_1) , onde $p_1 = 1 - p_0$. Usando a Figura 2.8 e considerando as condições de teste mostradas. Antes do particionamento a distribuição de classe é $(0,5, 0,5)$, pois os dados de treinamento tem quantidades iguais para cada classe. Particionando os dados utilizando o atributo *Gênero*, as distribuições de classes dos nós filhos são $(0,6, 0,4)$ e $(0,4, 0,6)$, que embora as classes não sejam mais igualmente distribuídas, os nós filhos contém elementos das duas classes. Particionando pelo atributo *Tipo de carro* tem-se como resultado partições mais puras.

As medidas desenvolvidas para a seleção dos melhores particionamentos em sua maioria baseiam-se no grau de impureza dos nós filhos. Quanto menor o grau de impureza, mais enviesado é a distribuição de classes.

Alguns critérios de partição:

- Índice de Gini:

$$1 - \sum_i p^2(i), \tag{2.17}$$

onde a soma é sobre todas as classes i no conjunto de dados e $p(i)$ é a fração observada das classes com classe i que chegam no nó. Um nó com apenas uma classe (nó puro) tem índice de Gini igual a zero, caso contrário um valor positivo. Portanto o índice de Gini é uma medida de impureza do nó.

- Entropia:

$$- \sum_i p(i) \log p(i) \tag{2.18}$$

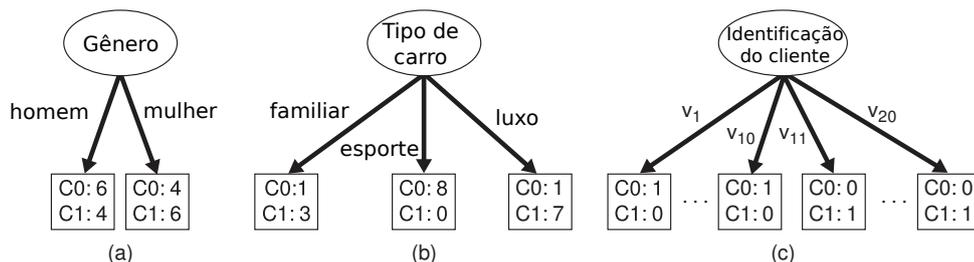


Figura 2.8: Partição binária e multicaminhos. Adaptado de Tan, Steinbach e Kumar (2006)

Um nó puro vai apresentar deviance 0, caso contrário será um valor positivo. $p(i)$ é definido da mesma forma que o índice anterior.

- Regra de Twoing:

Esta não é uma medida de pureza do nós, mas uma forma de medir como o nó deve fazer a partição dos dados. Tomando $L(i)$ como a fração dos elementos da classe i no nó da direita depois de particionado e $R(i)$ a fração dos elementos da classe i no nó da esquerda depois da partição. Utilizando o seguinte critério de partição para maximizar:

$$P(L)P(R) \left(\sum_i |L(i) - R(i)| \right)^2, \quad (2.19)$$

sendo $P(L)$ e $P(R)$ as frações dos dados que foram particionadas para a esquerda e direita respectivamente. Se o resultado da expressão for grande, a divisão tornou cada nó o mais puro possível, caso contrário, com um resultado pequeno, a divisão tornou os nós muito similares e conseqüentemente similares com o nó pai e essa partição não incrementou a pureza do nó.

- Erro:

O erro no nó é a fração das classes classificadas erradamente no nó. Sendo j a classe com o maior número de amostras de treinamento no nó, o erro é dado por:

$$1 - p(j) \quad (2.20)$$

2.3 Bases desbalanceadas

O problema de conjuntos de treinamentos desbalanceados é um problema bem conhecido. Ocorre quando uma das classes apresenta um número menor de pontos de treinamentos o que causa *overfitting* do classificador para a classe majoritária (ARISHOLM; BRIAND; JOHANNESSEN, 2010; HALL et al., 2012). Pode ser encontrado em diversas áreas: chamadas telefônicas fraudulentas (PROVOST; FAWCETT, 2001), gerenciamento de telecomunicações (EZAWA; SINGH; NORTON, 1996), classificação de texto (LEWIS; CATLETT, 1994), detecção de manchas de óleo em imagens de satélite (KUBAT; HOLTE; MATWIN, 1998).

Diversas soluções são propostas para esse problema, uma delas é subamostrar a classe majoritária, deixando-a com a mesma quantidade da classe minoritária ou próximo disso (KUBAT; HOLTE; MATWIN, 1998; JAPKOWICZ, 2000; DOMINGOS, 1999). O problema com esse método é escolher os dados apropriados de forma que não haja perda na informação a respeito daquela classe. Pode-se também fazer uma sobreamostragem dos

dados da classe minoritária pela replicação dos dados, fazendo com que seu número seja igual ao da classe com maior número de elementos (LING et al., 1998; JAPKOWICZ, 2000; DOMINGOS, 1999).

Em uma outra abordagem a classe minoritária é sobreamostrada por meio do algoritmo SMOTE (CHAWLA et al., 2002) que cria instâncias da classe minoritária, para cada amostra da classe com menos elementos uma nova amostra é criada na linha de ligação com o K-vizinho mais próximo (KNN).

As amostras sintéticas são geradas da seguinte forma: Fazendo a diferença do vetor de característica e o vizinho mais próximo. Multiplicando essa diferença por um número aleatório entre 0 e 1, então somando ao vetor característica considerado. Este método força a região de decisão da classe minoritária a se tornar mais geral.

Considere uma amostra (6,4) e sendo (4,3) seu vizinho mais próximo.

(6,4) é uma amostra na qual k-vizinhos próximos são identificados.

(4,3) é um de seus k-vizinhos próximos

Fazendo

$$f1_1=6 \quad f2_1 = 4 \quad f2_1-f1_1 = -2$$

$$f1_2=4 \quad f2_2 = 3 \quad f2_2-f1_2 = -1$$

As novas amostras serão geradas por

$$(f1', f2') = (6,4) + \text{rand}(0-1) * (-2, -1)$$

rand(0-1) gera um número aleatório entre 0 e 1.

Tabela 2.2: Exemplo da geração de dados utilizando SMOTE. Adaptada de Chawla et al. (2002)

2.4 Misturas

Experimentos com misturas são utilizados em diversas áreas de desenvolvimento e melhoria de produtos. Nestes experimentos, dois ou mais ingredientes (componentes) são misturados em diferentes proporções para formar um produto final. As características de qualidade do produto final são registradas para cada mistura e verificado se elas variam. Como definição: problemas com mistura as respostas medidas dependem somente das proporções dos componentes presentes na mistura e não na quantidade de mistura (CORNELL, 2002).

Tomando q como o número de componentes de uma mistura e representando a proporção do i -ésimo componente por x_i , então (CORNELL, 2002):

$$x_i \geq 0, \quad i = 1, 2, \dots, q \quad (2.21)$$

e

$$\sum_{i=1}^q x_i = x_1 + x_2 + \dots + x_q = 1 \quad (2.22)$$

De acordo com a Equação 2.22, a soma das proporções dos componentes é a unidade. Essa condição é uma restrição fundamental associada às proporções que compõem a mistura.

Satisfazendo as restrições das Equações 2.21 e 2.22 significa somente que a composição da mistura será formada pela adição de quantidades não-negativas. Uma proporção individual x_i deve ser igual a um, a mistura terá um único componente. É chamada de mistura pura ou mistura de único componente. Misturas de um único componente são usadas como um padrão quando misturas de vários componentes são comparadas.

As variáveis x_i são linearmente dependentes, de modo que a mudança na proporção de um componente da mistura ocasionará a mudança em pelo menos um dos demais componentes.

A descrição geométrica do espaço contendo componentes q consiste de todos os pontos dentro das fronteiras (vértices, bordas, lados, etc.) de um espaço *simplex*¹ regular de dimensão $q - 1$. Para $q = 2$ componentes a Equação 2.22 é reduzida para $x_1 + x_2 = 1$, representada pela reta $x_2 = 1 - x_1$. Todas as possíveis misturas para dois componentes correspondem a algum ponto sobre esta reta (Figura 2.9a). Com três componentes ($q = 3$), o espaço simplex é um triângulo equilátero (Figura 2.9b) e para $q = 4$ o espaço simplex é um tetraedro (Figura 2.9c).

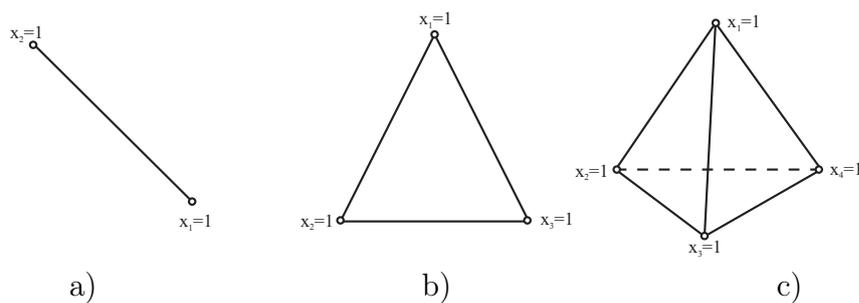


Figura 2.9: Regiões do espaço simplex para misturas de 2, 3 e 4 componentes. Adaptado de Cornell (2002)

¹Configuração espacial de n dimensões determinada por $n + 1$ pontos em um espaço de dimensão igual ou maior que n .

2.4.1 Mistura de dois componentes

Nas misturas de biodieseis de diferentes matérias-primas os valores dos ésteres dependem somente das suas proporções nos componentes, para encontrar a nova proporção de um dado éster em uma mistura pode-se usar a expressão:

$$y = \beta_1 x_1 + \beta_2 x_2 \quad (2.23)$$

onde β_1 e β_2 são as porcentagens do dado éster em cada uma das amostras e x_1 e x_2 são as novas proporções na mistura, com a restrição $x_1 + x_2 = 1$.

2.4.2 Mistura de três componentes

Uma mistura com três componentes deve ser obedecida a restrição $x_1 + x_2 + x_3 = 1$. Assim para uma mistura de três componentes, para saber a nova concentração de um dado éster usa-se a equação:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (2.24)$$

onde β_1 , β_2 e β_3 são as porcentagens do dado éster em cada uma das amostras e x_1 , x_2 e x_3 são as novas proporções na mistura, com a restrição $x_1 + x_2 + x_3 = 1$.

Capítulo 3

Metodologia Proposta

3.1 Base de dados

A base de dados consiste de 98 amostras contendo 13 ésteres dos seguintes ácidos graxos: ácido caprílico (C8:0), ácido cáprico (C10:0), ácido láurico (C12:0), ácido mirístico (C14:0), ácido palmítico (C16:0), ácido esteárico (C18:0), ácido oleico (C18:1), ácido linoleico (C18:2), ácido linolênico (C18:3), ácido araquídico (C20:0), ácido gadoleico (C20:1), ácido erúcido (C22:1) e ácido ricinoleico (C18:1 OH). Esta notação de ésteres (CX:Y), X indica a quantidade de átomos de carbono e Y corresponde ao número de insaturações presentes na moléculas. A base é formada em parte por amostras de biodieles elaborada pelo Laboratório de Análise e Pesquisa e Química Analítica de Petróleo (LAPQAP/UFMA) e a outra obtida a partir da literatura (BARRADAS FILHO et al., 2015; PARK et al., 2008; KIVEVELE; MBARAWA, 2011; BALAT, 2011; MOSER, 2011; MARTÍN et al., 2010; NOGUEIRA et al., 2010; SARIN et al., 2007; DANTAS, 2010; BARRADAS FILHO, 2015; CANDEIA, 2008; SANTOS, 2008; PADHI, 2010)

Na Figura 3.1a são mostradas as frequências dos ésteres metílicos de ácidos graxos (FAME, do inglês *fatty acid methyl ester*) na base de dados utilizada. Os ésteres mais encontrados foram C16:0, C18:0, C18:1 presentes em todas as amostras e C18:2 em 98,98%, C18:3 em 84,69% e C20:0 contido em 75,51%, mostrando consistência outros trabalhos (HOEKMAN et al., 2012; MENG; JIA; WANG, 2014). Alguns ésteres que aparecem em menor frequência na base de dados podem ser majoritários para algumas matérias-primas, como é o caso do C12:0 que corresponde a 52,51% do biodiesel de tucum, 41% no babaçu (SILVA et al., 2008) ou o C18:1 OH e chega a corresponder a 90,3% do biodiesel de mamona (DEMIRBAS, 2008; MARTÍN et al., 2010; BARRADAS FILHO et al., 2015; DANTAS et al., 2011b). Para resolver o problema das bases desbalanceadas foi utilizado neste trabalho o algoritmo SMOTE (CHAWLA et al., 2002). Na Figura 3.1b pode-se

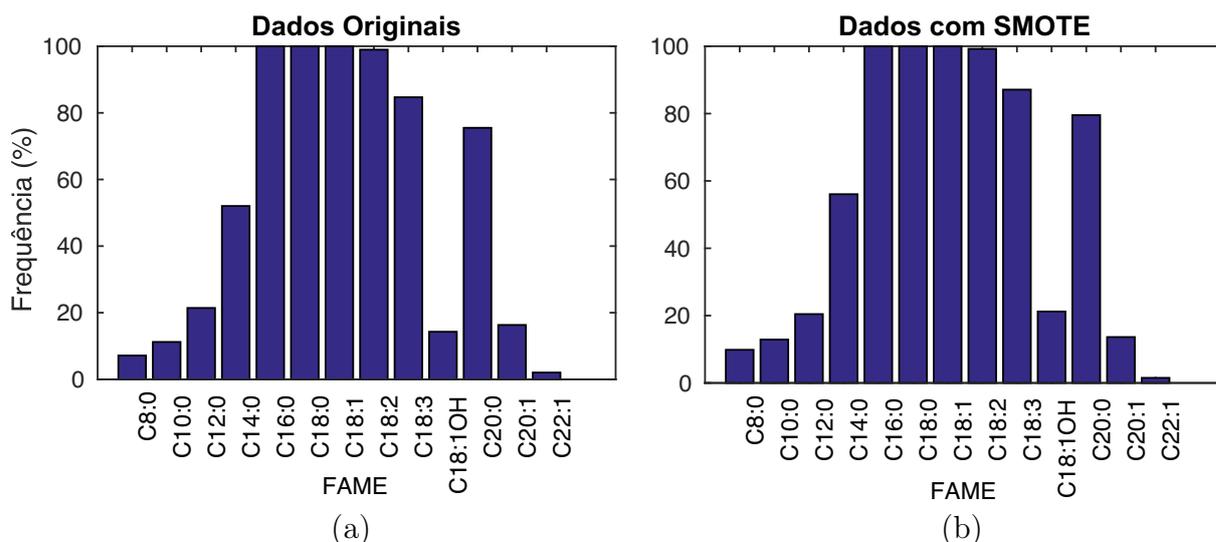


Figura 3.1: Frequência de cada éster para as amostras de biodiesel da base de dados utilizada. (a) Dados originais (b) Dados após o algoritmo SMOTE. Adaptada de (BARRADAS FILHO et al., 2015)

verificar a distribuição dos ésteres para a nova base de dados gerada, o conjunto original consistia de 66 amostras conformes e 26 amostras não-conformes para propriedade de viscosidade da norma EN 14214. Após SMOTE as duas classes apresentam o mesmo número de amostras (66) e como observado na figura a nova base de dados mantém a distribuição original. A base é composta de diferentes matérias-primas de origem vegetal e animal, como observado na Figura 3.2.

Foram estudadas quatro propriedades: viscosidade, índice de iodo, estabilidade oxidativa e massa específica com as amostras então rotuladas em conforme e não-conforme de acordo com cada uma das especificações utilizadas. Na Tabela 3.1 são mostradas algumas informações sobre a base de dados. As faixas de variação para os quatro parâmetros bem como as especificações para as resoluções de qualidade de biodiesel: EN 14214 do *Comité Européen de Normalisation* (CEN), D6751 da *American Society for Testing and Materials* (ASTM) e RANP 45/2014 da Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP). As amostras visualizadas em 3 dimensões com a redução em PCA são mostradas na Figura 3.3, rotuladas por propriedade e especificação.

Parâmetro	Número de Amostras	Unidade	Faixa de variação nas amostras		Especificação		
			Mínimo	Máximo	EN 14214:2014	ASTM D6751-15	RANP 45/2014
Viscosidade	98	mm^2/s	2,73	15,25	3,5-5,0	1,9-6,0	3,0-6,0
Índice de iodo	98	$g/100g$	14,67	171,50	120 (min)	-	-
Período de Indução	98	h	0,20	35,50	8 (min)	3 (min)	8 (min)
Massa específica	98	kg/m^3	807,30	915,70	860-900	-	850-900

Tabela 3.1: Amostras de biodiesel utilizadas e especificações adotadas para a verificação dos parâmetros de qualidade. Adaptada de Barradas Filho et al. (2015)

Das 98 amostras da base de dados foram utilizadas 92 para a etapa de treinamento dos classificadores e ANN e 6 amostras foram utilizadas para o teste de verificação final

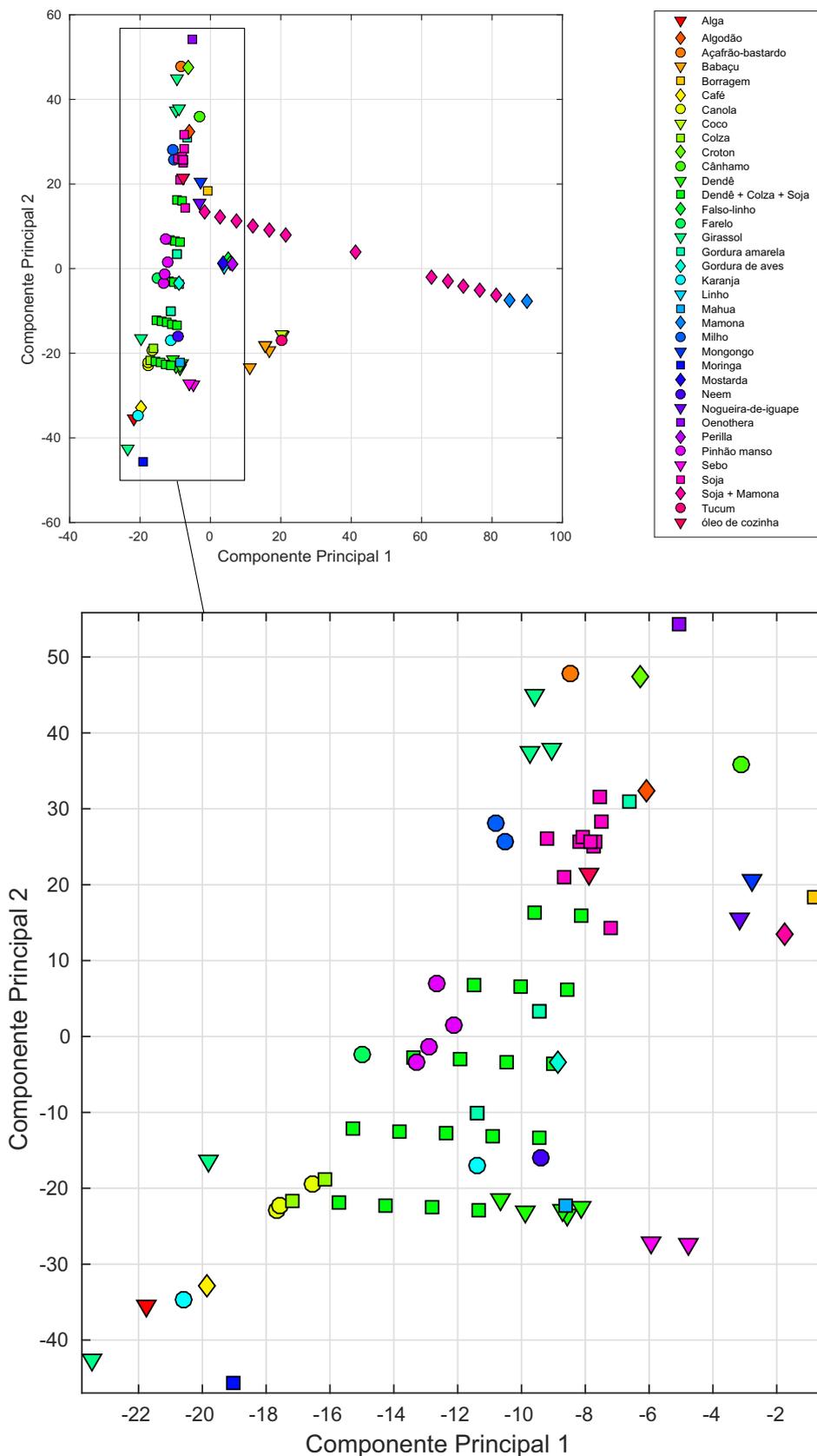


Figura 3.2: Projeção das 98 amostras para visualização em 2 dimensões utilizando PCA agrupadas por matéria-prima.

com 3 amostras conformes e 3 amostras não-conformes de acordo com as 3 especificações adotadas.

3.2 Normalizar/Padronizar/Escalonar

A normalização de um vetor significa dividi-lo pela sua norma, geralmente a norma euclidiana é utilizada. Para problemas como o algoritmo K -NN, outras normas também podem ser utilizadas: norma L1, distância Manhattan,

$$\mathbf{x}' = \frac{\mathbf{x}}{\|\mathbf{x}\|} \quad (3.1)$$

onde \mathbf{x} é o vetor original, \mathbf{x}' o vetor normalizado.

A padronização do vetor significa subtrair uma medida de localização e dividir por uma medida de escala. Como exemplo, para um vetor possuindo valores com uma distribuição gaussiana, uma forma de fazer isso seria subtrair a média e dividir o resultado pelo desvio padrão, obtendo assim um vetor com média zero e desvio padrão unitário.

$$\mathbf{z} = \frac{\mathbf{x} - \mu}{\sigma} \quad (3.2)$$

sendo \mathbf{x} o vetor original, μ e σ são sua média e desvio padrão, respectivamente.

O escalonamento é de grande importância para os algoritmos de aprendizagem. A principal vantagem do escalonamento é evitar que características com faixas maiores de valores representem maior importância para a classificação que características com faixas de valores menores. Para SVM, uma outra vantagem da utilização do escalonamento é evitar erros de ordem numérica, uma vez que o cálculo de kernel depende do produto interno de vetores, grandes valores podem levar a problemas numéricos. Uma recomendação é utilizar o escalonamento de cada atributo de forma linear na faixa $[-1, +1]$ ou $[0, 1]$ (HSU; CHANG; LIN, 2010). A fórmula para o escalonamento é dada por:

$$\mathbf{x}' = \frac{\mathbf{x} - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \quad (3.3)$$

sendo \mathbf{x} o vetor original e \mathbf{x}' o vetor escalonado.

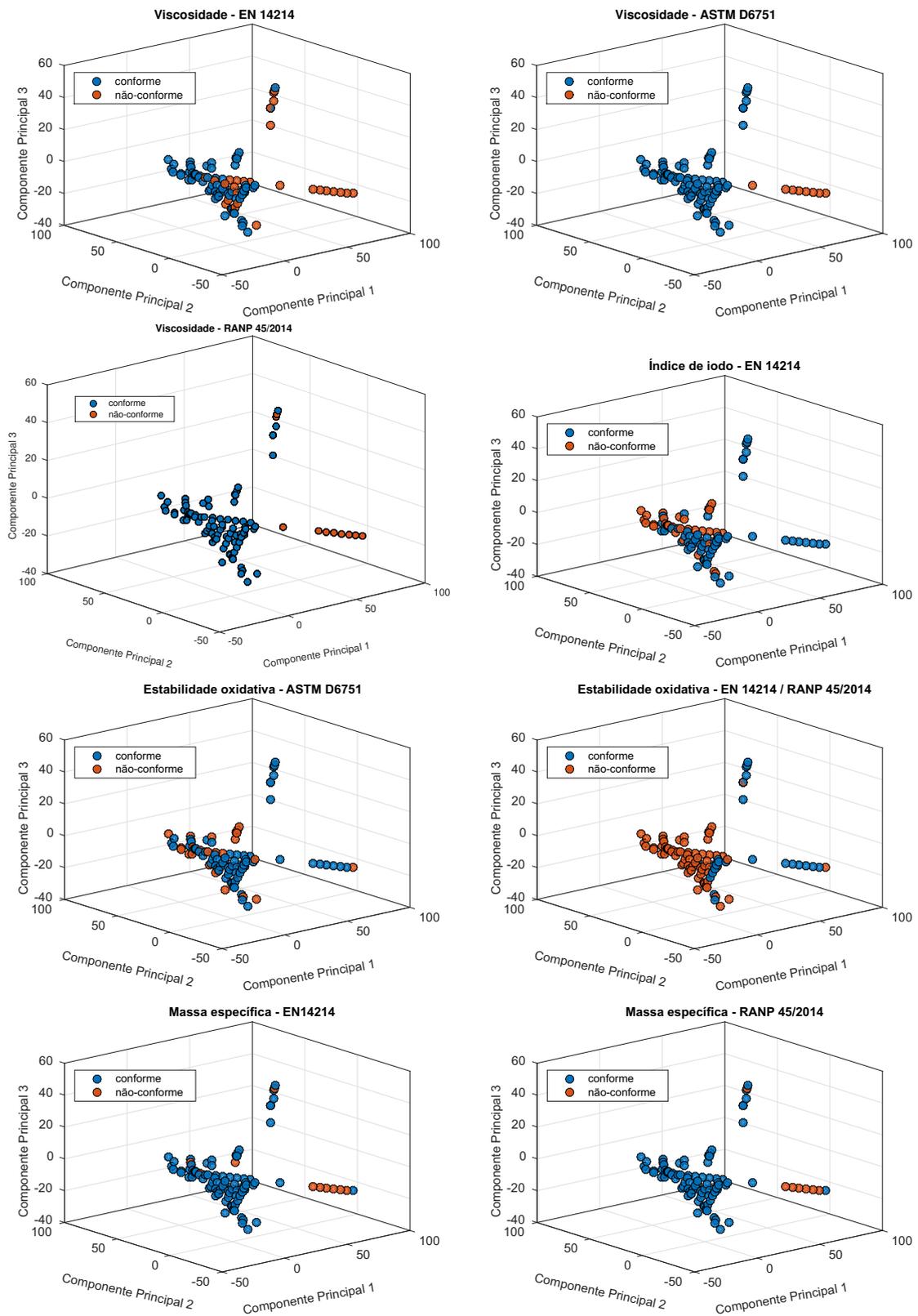


Figura 3.3: *Projeção das 98 amostras para visualização em 3 dimensões utilizando PCA para as propriedades de viscosidade, índice de iodo, estabilidade oxidativa e massa específica para as especificações EN 14214, ASTM D6751 e RANP 45/2014*

Kernel	Fórmula	Parâmetros do Kernel	Parâmetro C
Linear	$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$	-	
Polinomial	$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d$	$d = 2 - 6, \gamma = 2^{-15} - 2^3$	$C = 2^{-5} - 2^{15}$
Gaussiano	$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$	$\gamma = 2^{-15} - 2^3$	

Tabela 3.2: Parâmetros para a construção dos modelos de classificador SVM

3.3 Modelos dos classificadores

Para determinação dos parâmetros SVM os parâmetros C e λ devem ser escolhidos de forma apropriada para evitar *overfitting* do classificador. Os parâmetros são escolhidos com o melhor resultado de acurácia utilizando o método de *5-fold cross-validation*, que consiste em dividir o conjunto de treinamento em 5 subconjuntos de tamanho aproximado, um subconjunto é utilizada para o teste do classificador e os outros restantes são utilizado para o treinamento.

Para a implementação da seleção dos melhores parâmetros foi utilizada a biblioteca LIBSVM software package (CHANG; LIN, 2011), feita uma busca em grade (*grid-search*), com vários pares (C, γ) . Uma forma prática para a busca dos parâmetros é utilizar faixas exponencialmente crescentes. A faixa de valores utilizada para busca foi $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ e $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$ (HSU; CHANG; LIN, 2010). Os tipos de funções kernel utilizadas e as faixas de parâmetros estão na Tabela 3.2.

O classificador k -NN foi testado para diferentes tipos de medidas de distância e para diferentes valores de k . As distâncias e a faixa de k utilizada aparecem na Tabela 3.3. Os melhores valores de classificação foram utilizados como critério de escolha.

As árvores de decisão foram implementadas utilizando a Matlab toolbox, foram testados os critérios de decisão Gini's Diversity Index, Twoing rule e Deviance, para o algoritmo de escolha das melhores partições foram utilizados Exact, PullLeft, PCA e OVAbbyClass.

Distância	Fórmula	Número de vizinhos (K)
Euclidiana	$d_{st}^2 = (x_s - y_t)(x_s - y_t)'$	
Mahalanobis	$d_{st}^2 = (x_s - y_t)C^{-1}(x_s - y_t)'$ ^a	$K = 1-20$
Minkwoski	$d_{st} = \sqrt[p]{\sum_{j=1}^n x_{sj} - y_{tj} ^p}$	

Tabela 3.3: Parâmetros para construção dos modelos de classificação KNN

^aC é a matriz de covariância

Parâmetros	Especificações
Algoritmo de Convergência	Levenberg-Marquardt
Função de minimização do erro	Erro quadrático médio (MSE)
Número máximo de épocas	10000
Número de neurônios de entrada	13
Número de neurônios da camada oculta	5-30
Função de transferência da camada oculta	Tangente hiperbólica / Log-sigmoidal
Função de transferência da camada de saída	Função de transferência linear

Tabela 3.4: *Parâmetros de treinamento das Redes Neurais*

As redes neurais têm sido utilizadas em diversas pesquisas como um método alternativo para a predição de parâmetros de qualidade de biodiesel (YANG et al., 2002; LÔBO; FERREIRA; CRUZ, 2009; AL-HAMAMRE; AL-SALAYMEH, 2014). Dentre os diferentes tipos de arquiteturas de ANN, a literatura mostra a Multilayer Perceptron (MLP) como um modelo popular em publicações recentes (BARRADAS FILHO et al., 2015; ARUMUGAM; SRIRAM; SUBRAMANIAN, 2012).

Em coerência com a literatura, para o objetivo do presente trabalho foi implementado o modelo MLP para a classificação dos parâmetros de qualidade de biodiesel para as classes conforme e não-conforme. Para o treinamento das redes, foi utilizado o método de convergência Levenberg-Marquardt para minimizar o erro quadrático médio (MSE) com o número máximo de épocas definido como 10000. As funções tangente hiperbólica e log-sigmoidal foram testadas para ativação da camada oculta e a quantidade de neurônios ocultos foi variada de 5 a 30. A função de transferência linear foi definida como função de ativação da camada de saída (Tabela 3.4).

As redes neurais foram implementadas utilizando a arquitetura de MLP (do inglês, *multilayer perceptron*), com três camadas (entrada, camada oculta e saída). Cada rede foi treinada utilizando o método de *back-propagation*, sendo otimizado o número de neurônios na camada oculta e a função de transferência. Os parâmetros utilizados no treinamento podem ser vistos na Tabela 3.4. As redes neurais otimizadas utilizadas nos resultados estão apresentadas na Tabela 3.5.

Após a otimização dos quatro métodos de classificação direta (TREE, KNN, SVM e ANN) pelas etapas de treinamento e validação interna, as classificações das seis amostras de validação externa foram utilizadas para avaliar e comparar os métodos otimizados. Para cada parâmetro e para cada norma, um método do presente trabalho foi selecionado e comparado a equações empíricas obtidas da literatura com base nos resultados para as

amostras de validação externa. Esta etapa objetiva estabelecer um comparativo entre a classificação direta e a classificação indireta (predição do valor da propriedade seguida de classificação).

No trabalho de [Barradas Filho \(2015\)](#), o autor desenvolveu equações por MLR e comparou a algumas equações da literatura para prever viscosidade, estabilidade oxidativa (período de indução) e índice de iodo de biodiesel. As duas equações com menores erros, para cada uma dessas três propriedades, foram utilizadas neste trabalho. Para a propriedade massa específica, duas equações foram selecionadas de outros trabalhos da literatura e todas as equações podem ser conferidas na Tabela 3.6.

Para comparar de forma satisfatória os modelos dos classificadores, evitando a dependência dos dados de treinamento, o método de validação k-fold ($k = 5$) foi utilizado. Para avaliar os algoritmos de classificação de forma eficiente, o procedimento foi repetido 5 vezes e a média dos resultados apresentados e os parâmetros dos classificadores otimizados a partir dos valores nas Tabelas 3.2, 3.3 e 3.4.

Um fluxograma completo da metodologia adotada pode ser visto na Figura 3.4

3.4 Medidas de desempenho

Para medida de desempenho dos classificadores foram utilizadas as medidas de acurácia (ACC), sensibilidade (SENS) e especificidade (SPEC). A acurácia mostra a concordância entre os resultados do classificador e os valores de referência. É a medida da proporção de todos os resultados corretamente classificados pelo total de amostras testadas. Esta medida é semelhante à utilizada por ([BALABIN; SAFIEVA; LOMAKINA, 2010](#); [BALABIN; SAFIEVA, 2011](#)).

A sensibilidade é a habilidade do classificador em detectar amostras conformes quando elas são realmente conformes.

Parâmetro	Topologia	Especificação	Função de Transf. Camada Oculta
Viscosidade	13-19-1	RANP 45/2014	Log-sigmoidal
	13-16-1	EN 14214:2014	Log-sigmoidal
	13-9-1	ASTM D6751-15	Log-sigmoidal
Estabilidade oxidativa	13-7-1	RANP 45/2014 e EN 14214:2014	Log-sigmoidal
	13-30-1	ASTM D6751-15	Log-sigmoidal
Massa específica	13-26-1	RANP 45/2014	Log-sigmoidal
	13-16-1	EN 14214:2014	Log-sigmoidal
Índice de iodo	13-27-1	EN 14214:2014	Log-sigmoidal

Tabela 3.5: Resultados da otimização das redes neurais

A especificidade representa a habilidade do classificador identificar corretamente uma amostra não-conforme, quando ela é não-conforme.

$$\text{Acurácia} = \frac{TN + TP}{TN + TP + FN + FP} \quad (3.4)$$

$$\text{Sensibilidade} = \frac{TP}{TP + FN} \quad (3.5)$$

$$\text{Especificidade} = \frac{TN}{TN + FP} \quad (3.6)$$

sendo TP é o número de amostras conformes que foram corretamente identificadas, TN é o número de amostras não-conformes que foram corretamente identificadas, FP é o número de amostras conformes que foram identificadas como não-conformes e FN é o número de amostras não-conformes que foram identificadas como conformes.

Parâmetro	Notação	Equação do Modelo	Variáveis
Viscosidade	Modelo 1 (HONG; JEON; LEE, 2014)	$KV_{pr} = \sum x_i KV_i$	KV_{pr} = viscosidade do biodiesel x_i = fração da massa do i-ésimo éster metílico KV_i = viscosidade do i-ésimo éster metílico puro
	Modelo 2 (BARRADAS FILHO et al., 2015)	$V = 2.7251 + (0.12579 \times C8 : 0) +$ $+ (-0.38892 \times C10 : 0) + (0.00446 \times C12 : 0) +$ $+ (0.06446 \times C14 : 0) + (0.01904 \times C16 : 0) +$ $+ (0.00905 \times C18 : 0) + (0.02349 \times C18 : 1) +$ $+ (0.01363 \times C18 : 2) + (0.00745 \times C18 : 3) +$ $+ (0.1228 \times C18 : 1 OH) + (0.06315 \times C20 : 0) +$ $+ (0.00597 \times C20 : 1) + (0.03667 \times C22 : 1)$	V = viscosidade do biodiesel $C8:0$ = % (m/m) éster metílico do ácido caprílico $C10:0$ = % (m/m) éster metílico do ácido cáprico $C12:0$ = % (m/m) éster metílico do ácido láurico $C14:0$ = % (m/m) éster metílico do ácido mirístico $C16:0$ = % (m/m) éster metílico do ácido palmítico $C18:0$ = % (m/m) éster metílico do ácido esteárico $C18:1$ = % (m/m) éster metílico do ácido oleico $C18:2$ = % (m/m) éster metílico do ácido linoleico $C18:3$ = % (m/m) éster metílico do ácido linolênico $C18:1 OH$ = % (m/m) éster metílico do ácido riconoleico $C20:0$ = % (m/m) éster metílico do ácido araquídico $C20:1$ = % (m/m) éster metílico do ácido gadoleico $C22:1$ = % (m/m) éster metílico do ácido erúico
Índice de Iodo	Modelo 3 (WANG et al., 2012)	$IV = 0.6683 \times DU + 25.0364$ $DU = (C_n:1) + 2 \times (C_n:2) +$ $+ 3 \times (C_n:3) + 4 \times (C_n:4)$	IV = índice de iodo biodiesel DU = grau de insaturação $C_n:1$ = % em massa de $CX:1$ $C_n:2$ = % em massa de $CX:2$ $C_n:3$ = % em massa de $CX:3$ $C_n:4$ = % em massa de $CX:4$
	Modelo 4 (BARRADAS FILHO et al., 2015)	$IV = -6.9512 + (-1.01098 \times C8 : 0) +$ $+ (4.71488 \times C10 : 0) + (0.20634 \times C12 : 0) +$ $+ (-0.58867 \times C14 : 0) + (-0.1825 \times C16 : 0) +$ $+ (1.01984 \times C18 : 0) + (1.14504 \times C18 : 1) +$ $+ (1.72482 \times C18 : 2) + (1.99429 \times C18 : 3) +$ $+ (1.04122 \times C18 : 1OH) + (3.00665 \times C20 : 0) +$ $+ (2.11614 \times C20 : 1) + (-0.13725 \times C22 : 1)$	IV = índice de iodo biodiesel $C8:0$ = % (m/m) éster metílico do ácido caprílico $C10:0$ = % (m/m) éster metílico do ácido cáprico $C12:0$ = % (m/m) éster metílico do ácido láurico $C14:0$ = % (m/m) éster metílico do ácido mirístico $C16:0$ = % (m/m) éster metílico do ácido palmítico $C18:0$ = % (m/m) éster metílico do ácido esteárico $C18:1$ = % (m/m) éster metílico do ácido oleico $C18:2$ = % (m/m) éster metílico do ácido linoleico $C18:3$ = % (m/m) éster metílico do ácido linolênico $C18:1 OH$ = % (m/m) éster metílico do ácido riconoleico $C20:0$ = % (m/m) éster metílico do ácido araquídico $C20:1$ = % (m/m) éster metílico do ácido gadoleico $C22:1$ = % (m/m) éster metílico do ácido erúico
Período de Indução	Modelo 5 (PARK et al., 2008)	$Y = \frac{117.9295}{X} + 2.5905$	Y = período de indução do biodiesel X = soma do percentual de massa dos ésteres metílicos linoleico e linolênico
	Modelo 6 (BARRADAS FILHO et al., 2015)	$IP = 6.1924 + (0.85242 \times C8:0) +$ $+ (-1.31462 \times C10:0) + (0.24307 \times C12:0) +$ $+ (0.23174 \times C14:0) + (0.15171 \times C16:0) +$ $+ (-0.13884 \times C18:0) + (0.01139 \times C18:1) +$ $+ (-0.07447 \times C18:2) + (-0.06931 \times C18:3) +$ $+ (0.07772 \times C18:1 OH) + (-0.67244 \times C20:0) +$ $+ (-0.05735 \times C20:1) + (-0.07329 \times C22:1)$	IP = período de indução do biodiesel $C8:0$ = % (m/m) éster metílico do ácido caprílico $C10:0$ = % (m/m) éster metílico do ácido cáprico $C12:0$ = % (m/m) éster metílico do ácido láurico $C14:0$ = % (m/m) éster metílico do ácido mirístico $C16:0$ = % (m/m) éster metílico do ácido palmítico $C18:0$ = % (m/m) éster metílico do ácido esteárico $C18:1$ = % (m/m) éster metílico do ácido oleico $C18:2$ = % (m/m) éster metílico do ácido linoleico $C18:3$ = % (m/m) éster metílico do ácido linolênico $C18:1 OH$ = % (m/m) éster metílico do ácido riconoleico $C20:0$ = % (m/m) éster metílico do ácido araquídico $C20:1$ = % (m/m) éster metílico do ácido gadoleico $C22:1$ = % (m/m) éster metílico do ácido erúico
Massa específica	Modelo 7 (AGARWAL; SINGH; CHAURASIA, 2010)	$Y = (0,91613 - 0,00045x_1 - 0,000794x_2 -$ $- 0,00061x_3 - 0,00033x_4 - 0,0005x_5 -$ $- 0,00039x_6) \times 1000$	Y = massa específica do biodiesel x_1 = % (m/m) éster metílico do ácido palmítico x_2 = % (m/m) éster metílico do ácido esteárico x_3 = % (m/m) éster metílico do ácido oleico x_4 = % (m/m) éster metílico do ácido linoleico x_5 = % (m/m) éster metílico do ácido linolênico x_6 = % (m/m) éster metílico do ácido erúico
	Modelo 8 (RAMIREZ-VERDUZCO, 2013)	$\rho_i = 1,069 + \frac{3,575}{M_i} + 0,0113 \times$ $\times N - 7,41 \times 10^{-4} \times T$	ρ_i = massa específica do i-ésimo éster metílico puro M_i = peso molecular do i-ésimo éster metílico N = número de ligações duplas do i-ésimo éster metílico T = temperatura em Kelvin ($T = 293,15K$)

Tabela 3.6: Modelos lineares utilizados para predição das propriedades de viscosidade, índice de iodo, estabilidade oxidativa e massa específica (BARRADAS FILHO, 2015)



Figura 3.4: Fluxograma com as principais partes da metodologia utilizada no trabalho.

Capítulo 4

Resultados e Discussões

4.1 Resultados do algoritmo SMOTE

Para cada parâmetro de qualidade de biodiesel e para as especificações das diferentes normas, as 92 amostras destinadas ao desenvolvimento das técnicas de classificação foram submetidas ao SMOTE. Como há no total nove especificações (para os quatro parâmetros) pelas três normas, e apenas duas especificações são coincidentes (limite mínimo de 8 h para a estabilidade oxidativa pela EN 14214 e pela RANP 45/2014), foram obtidas oito diferentes bases de dados após a etapa de *over-sampling*.

Ressalta-se que apenas 92 amostras foram utilizadas como dados de entrada para o incremento de amostras sintéticas nas bases de dados desbalanceadas, pois 6 amostras (3 conformes e 3 não-conformes) para cada tipo de base de dados foram separadas para a etapa de validação externa. Na tabela 4.1 podem ser observadas as informações da quantidade de amostras conformes e não-conformes nas bases de dados geradas pela aplicação do método SMOTE. As amostras utilizadas para teste podem ser visualizadas na Figura 4.1.

Parâmetros	EN14214		RANP 45/2014		ASTM D6751	
	Amostras conformes	Amostras não-conformes	Amostras conformes	Amostras não-conformes	Amostras conformes	Amostras não-conformes
Viscosidade	66	26 (+40)	85	7 (+78)	87	5 (+82)
Índice de Iodo	59	33 (+26)	-	-	-	-
Estabilidade oxidativa	22 (+48)	70	22 (+48)	70	71	21 (+50)
Massa específica	85	7 (+78)	88	4 (+84)	-	-

Tabela 4.1: Número de amostras para cada classe de acordo com as especificações das normas EN 14214, ASTM D6751-15 e RANP 45/2014 e o número de amostras sintéticas produzidas com o algoritmo SMOTE (entre parêntesis), excluindo as 6 amostras utilizadas para validação externa

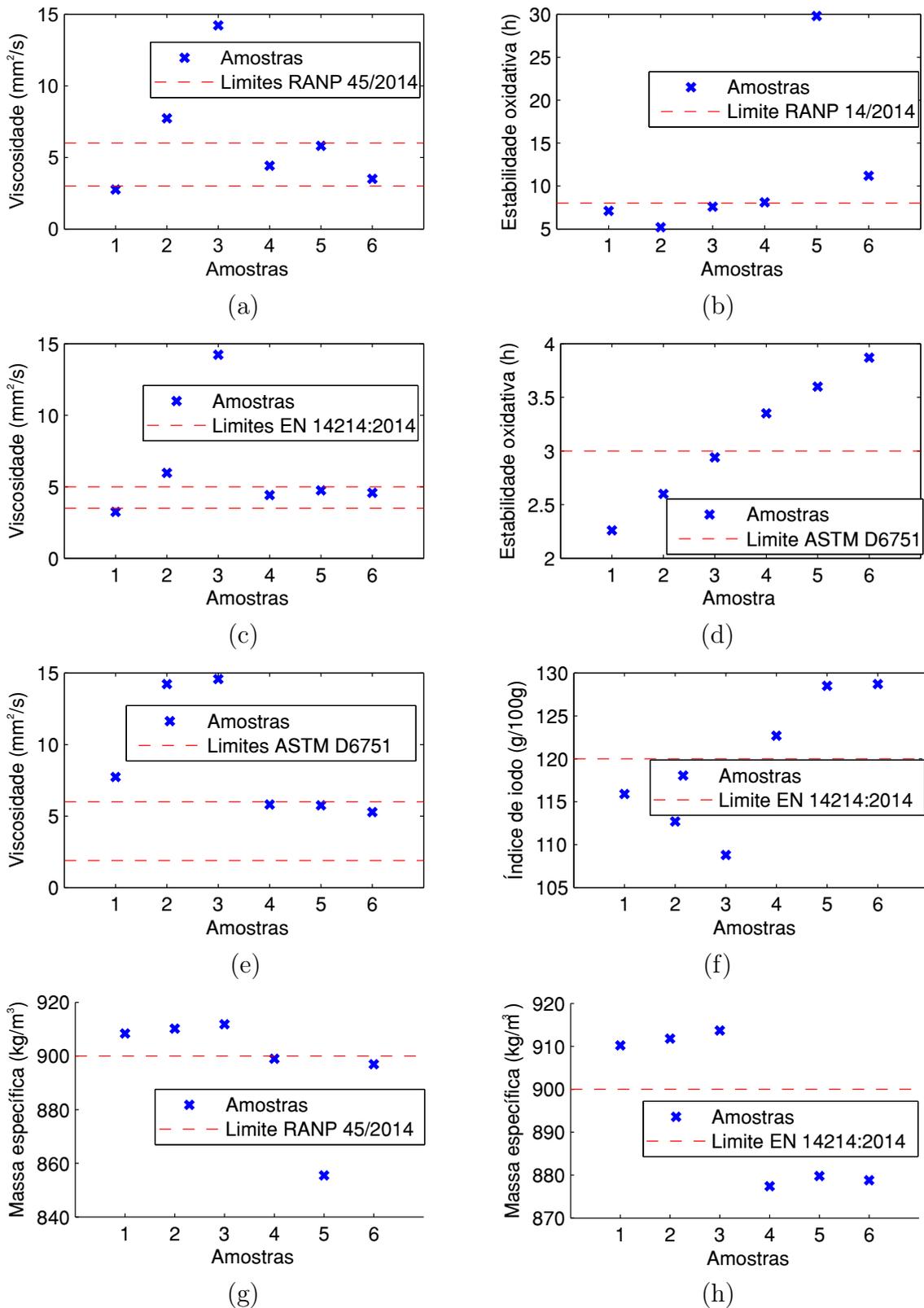


Figura 4.1: Amostras utilizadas para teste da propriedade: viscosidade (a) RANP 45/2014 (c) EN 14214:2014 e (e) ASTM D6751-15; estabilidade oxidativa (b) RANP 45/2014 e EN 14214:2014, (d) ASTM D6751-15; índice de iodo (f) (EN 14214:2014) e massa específica (g) RANP 45/2014 e (h) EN 14214:2014.

A norma EN 14214 é a mais restritiva em todos os parâmetros, enquanto que a ASTM D6751 é a menos restritiva para viscosidade e estabilidade oxidativa. As classificações das propriedades viscosidade, índice de iodo e densidade geraram a predominância de amostras conformes nos bancos de dados originais.

A viscosidade possui limites distintos para as três normas estudadas, de acordo com a Tabela 3.1, sendo os limites estabelecidos pela norma EN 14214 os mais restritivos e, portanto, originando mais amostras não-conformes em comparação às classificações pelas demais normas. O banco de dados original com classificação da viscosidade pela ASTM D6751 foi o que teve menos amostras não-conformes (apenas 5) e necessitou de mais amostras sintéticas para igualar com a quantidade de amostras conformes, devido aos limites dessa norma serem os menos restritivos.

A maioria dos casos de não-conformidade por excedência do limite máximo mais restritivo de viscosidade ($5,0 \text{ mm}^2/\text{s}$) ocorreram em amostras de misturas com pelo menos 15% de biodiesel de mamona (cerca de 13,5% de C18:1 OH). O éster C18:1 OH compõe aproximadamente 90% do biodiesel de mamona e, devido ao grupo hidroxila na molécula, eleva consideravelmente a viscosidade (BARRADAS FILHO et al., 2015; KNOTHE; STEIDLEY, 2005).

A conformidade das amostras quanto à massa específica apresentou comportamento semelhante à viscosidade, pois a classificação pela norma EN 14214, mais restritiva, gerou mais amostras não-conformes para o banco de dados original em comparação com a RANP 45/2014, que teve apenas 4 amostras não-conformes e precisou de mais amostras sintéticas. Para a propriedade índice de iodo, que só apresenta limite pela norma EN 14214, só havia 33 amostras não-conformes, necessitando de 26 amostras sintéticas para o balanceamento das classes.

A estabilidade oxidativa é a propriedade com maior diferença entre os limites estabelecidos pelas normas estudadas. Nas classificações das amostras pela norma EN 14214, a mais restritiva, predominaram as amostras não-conformes e o conjunto de amostras conformes, originalmente com 22 amostras, necessitou de 48 amostras sintéticas. Por outro lado, nas classificações pela ASTM D6751 predominaram as amostras conformes e o conjunto de amostras não-conformes necessitou de 50 amostras sintéticas.

Todas as amostras em conformidade com o limite de estabilidade oxidativa estabelecido pela EN 14214 estavam também conformes em relação ao índice de iodo, reforçando a relação direta entre esses dois parâmetros de qualidade (KEGL; KEGL; PEHAN, 2013; SARIN, 2012).

Testes preliminares realizados com os dados originais (sem SMOTE) confirmaram um comportamento já esperado atrelado ao desbalanceamento entre as classes. Para classi-

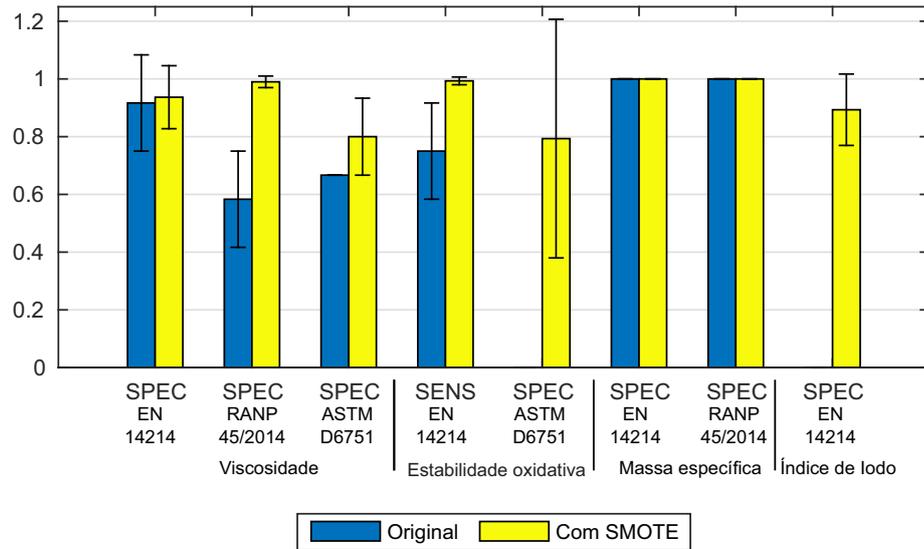


Figura 4.2: Médias dos resultados obtidos para SVM, KNN, TREE e ANN para o conjunto de amostras externo: comparação entre os dados originais e com a aplicação do algoritmo SMOTE

ficações cujos bancos de dados continham menor quantidade de amostras não-conformes, foram obtidas maiores taxas de falso positivo e menores taxas de verdadeiros negativos, afetando diretamente a especificidade. Por outro lado, em classificações com menor quantidade de amostras conformes, a taxa de falsos negativos foi elevada, reduzindo a sensibilidade dos métodos.

A Figura 4.2 apresenta um comparativo entre os resultados obtidos para as amostras de validação externa utilizando os dados originais e os dados com SMOTE. Os resultados dos parâmetros (sensibilidade e especificidade) que constam nos gráficos correspondem às médias dos resultados dos quatro métodos de classificação, e as barras correspondem ao desvio padrão.

Para propriedades e normas cujos dados de treinamento e validação interna continham menos amostras não-conformes, a comparação entre os resultados na Figura 4.2 foi feita apenas para a especificidade (parâmetro diretamente afetado por esse tipo de desbalanceamento). Unicamente para a estabilidade oxidativa pela norma EN 14214, cujos dados originais apresentaram menores quantidades de amostras conformes, os resultados para as amostras de teste para os dados originais e os dados com SMOTE foram comparados em termos da sensibilidade.

Para a classificação pela viscosidade, os modelos considerando as normas ASTM D6751 e RANP 45/2014 seriam os mais prejudicados pelo desbalanceamento das classes no banco de dados original, aumentando a taxa de falsos positivos e diminuindo a especificidade dos métodos. Usando os dados originais para o ajuste, todos os modelos tiveram em média especificidade de 0,5833 para RANP 45/2014 e 0,6667 para ASTM D6751, enquanto que o uso do SMOTE aumentou para 0,9900 e 0,8000 as especificidades

para RANP 45/2014 e ASTM D6751, respectivamente. O ajuste dos modelos para classificação da viscosidade pela norma EN 14214 quase não foi afetado pelo uso do SMOTE pois o banco de dados original já continha muitas amostras não-conformes, reduzindo o desbalanceamento entre as classes.

Ainda que o banco de dados do índice de iodo tenha apresentado o menor desbalanceamento entre as classes, a tendência gerada pela desproporcionalidade foi a mais evidente, em comparação com os outros bancos de dados originais. O ajuste dos modelos para classificação dessa propriedade, pela norma EN 14214, gerou especificidade média de 0,8933 usando o SMOTE, enquanto que todos os classificadores tiveram especificidade 0 com ajuste pelos dados originais.

Contrariamente ao banco de dados do índice de iodo, o banco de dados de classificação da densidade pela RANP 45/2014 teve o maior desbalanceamento entre as classes, e tanto para essa norma quanto para a EN 14214 não houve diferenças entre as especificidades dos ajustes pelos bancos de dados originais ou com SMOTE. Isso se deve ao fato de que as amostras do conjunto não-conforme sejam em sua maioria amostras de soja, mamona e misturas destas duas. É também possível que para essa propriedade tenha ocorrido *overfitting* por conta da pouca variedade das amostras, necessitando portanto de mais dados para treinamento.

A classificação da estabilidade oxidativa pela EN 14214 foi avaliada pela sensibilidade devido à presença majoritária de amostras não-conformes no banco de dados original. Utilizando SMOTE, quase todos os métodos de classificação tiveram sensibilidade máxima, gerando uma média de 0,9933. Sem SMOTE, a sensibilidade média era de 0,7500. Na classificação pela ASTM D6751, para a validação externa, a especificidade média foi de 0,7933 com o ajuste dos modelos pelos dados com SMOTE, enquanto que para o ajuste pelo banco de dados original nenhuma amostra não-conforme foi classificada corretamente, gerando uma especificidade média 0. Para a classificação pela ASTM D6751, na comparação entre os ajustes pelo dado original e dado com SMOTE, foi utilizado o parâmetro especificidade pois as amostras não-conformes constituíam a classe minoritária.

Visto que o ajuste dos métodos de classificação utilizando os bancos de dados com adição de amostras sintéticas pelo SMOTE apresentou resultados bem melhores do que usando os bancos de dados originais, para quase todas as propriedades, optou-se por utilizar os bancos de dados com SMOTE nas etapas de treinamento e de validação interna para que os modelos para classificação fossem mais aperfeiçoados, apesar do inevitável desbalanceamento que ocorreu em todos os bancos de dados originais.

4.2 Resultados dos modelos de classificadores

Os resultados de acurácia, sensibilidade e especificidade de cada método de classificação otimizado, para os três conjuntos de amostras (treinamento, validação interna e validação externa) e para as classificações dos parâmetros estudados de acordo com cada especificação são apresentados na Tabela 4.2. Os melhores resultados estão sombreados.

No presente trabalho, o principal critério utilizado para avaliar e comparar os métodos de classificação foi o desempenho nas amostras de validação externa, uma vez que estas simulam uma aplicação real dos métodos. Como critérios secundários, foram considerados a complexidade, o uso de memória e a velocidade de otimização dos métodos. A Figura 4.3 mostra os resultados de acurácia, sensibilidade e especificidade dos métodos para cada propriedade, apenas para as amostras de validação externa.

Na classificação da viscosidade pela norma EN 14214, os métodos SVM e KNN tiveram a maior acurácia para as amostras de validação externa (0,8667). Entretanto, enquanto que SVM teve maior sensibilidade (0,7467), KNN teve a máxima especificidade (1,000). Neste caso, é mais adequado optar pelo método com maior especificidade, pois garante uma taxa de verdadeiro negativo maior e, conseqüentemente, taxa de falso positivo menor.

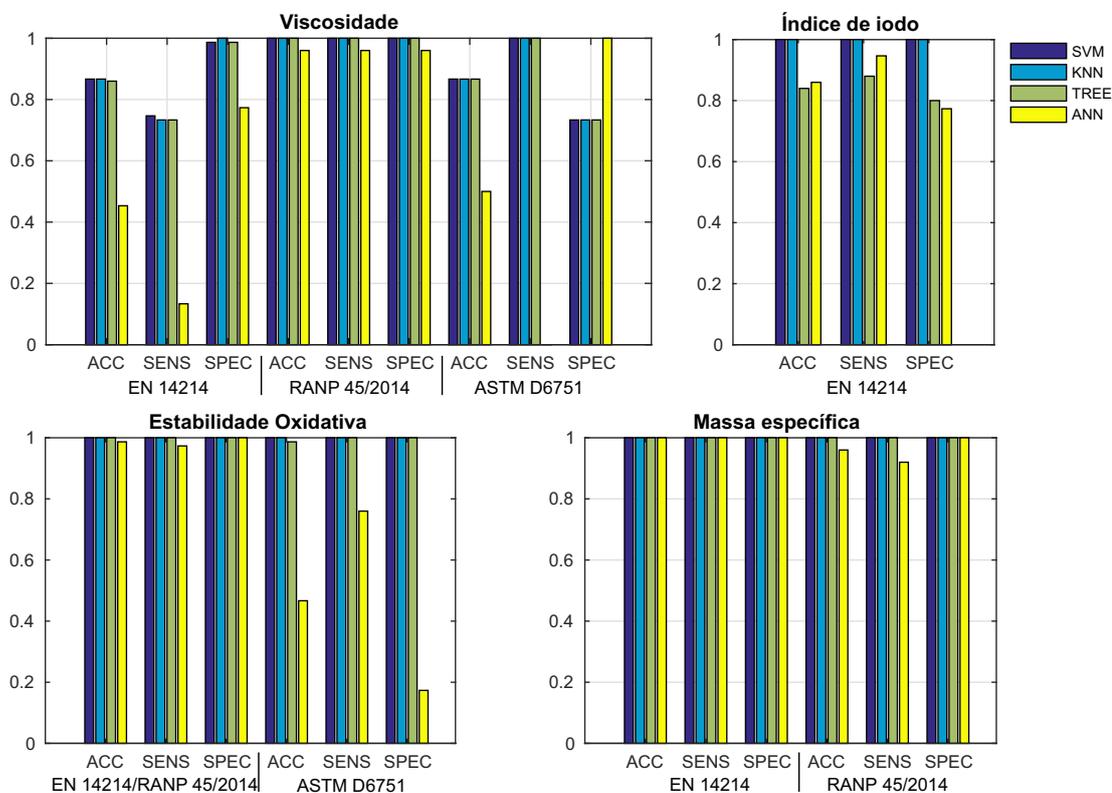


Figura 4.3: Comparação do desempenho entre os métodos SVM, KNN, TREE e ANN para classificação de amostras de biodiesel do conjunto de validação externo, para as propriedades de viscosidade, índice de iodo, estabilidade oxidativa e massa específica

Parâmetro	Norma	Método	Treinamento			Validação Interna			Validação Externa		
			ACC	SENS	SPEC	ACC	SENS	SPEC	ACC	SENS	SPEC
EN 14214		SVM	0,9498	0,9697	0,9302	0,8631	0,8685	0,8581	0,8667	0,7467	0,9867
		KNN	0,9936	0,9871	1,0000	0,8341	0,8300	0,8412	0,8667	0,7333	1,0000
		TREE	0,9596	0,9666	0,9527	0,8011	0,7823	0,8182	0,8600	0,7333	0,9867
		ANN	0,8755	0,9255	0,8264	0,8177	0,8568	0,7727	0,4533	0,1333	0,7733
RANP 45/2014		SVM	1,0000	1,0000	1,0000	0,9846	0,9856	0,9870	1,0000	1,0000	1,0000
		KNN	1,0000	1,0000	1,0000	0,9857	0,9787	0,9962	1,0000	1,0000	1,0000
		TREE	0,9989	1,0000	0,9978	0,9800	0,9720	0,9877	1,0000	1,0000	1,0000
		ANN	0,9961	0,9967	0,9955	0,9858	0,9899	0,9843	0,9600	0,9600	0,9600
ASTM D6751		SVM	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,8667	1,0000	0,7333
		KNN	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,8667	1,0000	0,7333
		TREE	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,8667	1,0000	0,7333
		ANN	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,5000	0,0000	1,0000
EN 14214		SVM	0,9940	1,0000	0,9878	0,8758	0,8955	0,8605	1,0000	1,0000	1,0000
		KNN	1,0000	1,0000	1,0000	0,8443	0,7952	0,8855	1,0000	1,0000	1,0000
		TREE	0,9243	0,9214	0,9276	0,8164	0,7976	0,8276	0,8400	0,8800	0,8000
		ANN	0,9384	0,9308	0,9469	0,8890	0,8598	0,9194	0,8600	0,9467	0,7733
RANP 45/2014		SVM	0,9881	0,9937	0,9821	0,8842	0,8928	0,8831	1,0000	1,0000	1,0000
		KNN	0,9946	1,0000	0,9891	0,8800	0,9088	0,8542	1,0000	1,0000	1,0000
		TREE	0,9810	0,9789	0,9828	0,9072	0,9014	0,9221	1,0000	1,0000	1,0000
		ANN	0,9586	0,9696	0,9480	0,9174	0,9282	0,9098	0,9867	0,9733	1,0000
ASTM D6751		SVM	0,9987	1,0000	0,9972	0,8315	0,8407	0,8284	1,0000	1,0000	1,0000
		KNN	1,0000	1,0000	1,0000	0,8005	0,8006	0,8032	1,0000	1,0000	1,0000
		TREE	0,9501	0,9354	0,9640	0,7834	0,7690	0,8097	0,9867	1,0000	1,0000
		ANN	0,8649	0,8560	0,8746	0,7994	0,7818	0,8153	0,4667	0,7600	0,1733
EN 14214		SVM	0,9679	0,9689	0,9669	0,9468	0,9441	0,9490	1,0000	1,0000	1,0000
		KNN	1,0000	1,0000	1,0000	0,9046	0,9292	0,8859	1,0000	1,0000	1,0000
		TREE	0,9778	0,9844	0,9713	0,8771	0,8949	0,8616	1,0000	1,0000	1,0000
		ANN	0,9544	0,9545	0,9548	0,9258	0,9457	0,9059	1,0000	1,0000	1,0000
RANP 45/2014		SVM	0,9894	0,9891	0,9896	0,9862	0,9840	0,9884	1,0000	1,0000	1,0000
		KNN	1,0000	1,0000	1,0000	0,9748	0,9602	0,9884	1,0000	1,0000	1,0000
		TREE	0,9962	1,0000	0,9924	0,9806	0,9674	0,9928	1,0000	1,0000	1,0000
		ANN	0,9896	0,9847	0,9946	0,9817	0,9723	0,9884	0,9600	0,9200	1,0000

Tabela 4.2: Resultados obtidos para SVM, KNN, TREE e ANN para classificação de viscosidade, índice de iodo, estabilidade oxidativa e massa específica para as normas EN 14214, ASTM D6751-15 e RANP 45/2014

Em termos de qualidade de combustível, a classificação de uma amostra não-conforme como conforme (falso positivo) pode ser mais nociva do que a classificação de uma amostra conforme como não-conforme (falso negativo), pois no primeiro caso poderia levar um biodiesel com viscosidade não-conforme a ser comercializado, causando danos para o motor. Desse modo, podemos selecionar o método KNN como o melhor para classificação da viscosidade pela EN 14214. As configurações do método KNN que alcançaram os melhores resultados em todas as três normas foram $k = 1$ e o método de busca baseado na distância euclidiana.

Considerando a classificação da viscosidade pelos limites da RANP 45/2014, os métodos SVM, KNN e TREE tiveram os mesmos resultados na etapa de validação externa: máximas acurácia, sensibilidade e especificidade. Algo semelhante aconteceu para a classificação pela norma ASTM D6751: SVM, KNN e TREE tiveram os mesmos resultados de acurácia (0,8667), sensibilidade (1,0000) e especificidade (0,7333). Em ambos os casos, o método selecionado foi o TREE, escolha motivada pela mínima complexidade entre os métodos estudados e baixo consumo de memória.

As configurações de TREE que apresentaram os melhores resultados foram Gini's Diversity Index para o critério de decisão e o algoritmo Exact para a escolha das melhores partições, para as classificações de viscosidade pelas normas ASTM D6751 e RANP 45/2014.

Na classificação das amostras de validação externa de acordo com o limite de índice de iodo da norma EN 14214, os métodos SVM e KNN apresentaram desempenho máximo nos três parâmetros avaliados (acurácia, sensibilidade e especificidade), o que indica que ambos os métodos classificaram corretamente todas as seis amostras.

Mais uma vez considerando o critério de complexidade do método, TREE foi selecionado como mais indicado para a classificação do índice de iodo, nas condições do presente trabalho, com o critério de decisão Gini's Diversity Index e o algoritmo Exact.

A classificação das amostras de validação externa para a estabilidade oxidativa, pelos limites das normas EN 14214 e RANP 45/2014, teve resultados muito bons para os quatro métodos. Isso pode ser justificado pela presença de muitas amostras com valores de estabilidade oxidativa bem próximos ao limite (8 h), favorecendo a melhor delimitação entre as classes.

Apesar dos quatro métodos terem apresentado bons resultados, três tiveram máxima acurácia, sensibilidade e especificidade para a estabilidade oxidativa: SVM, KNN e TREE. Pela menor complexidade, recomenda-se o uso do método TREE para classificação dessa propriedade de acordo com a norma EN 14214, com o algoritmo Exact e o critério de decisão Gini's Diversity Index.

Na classificação da estabilidade oxidativa pela ASTM D6751, foram obtidos bons resultados para três métodos, mas apenas SVM e KNN tiveram acurácia, sensibilidade e especificidade máximas. A classificação das amostras de validação externa por ANN teve o pior resultado de acurácia e especificidade dentre todos os parâmetros e normas, apesar da boa distribuição de amostras próximo do limite da ASTM D6751 (3 h).

O classificador KNN otimizado, selecionado para a classificação da estabilidade oxidativa pela norma ASTM D6751, foi baseado na distância euclidiana, com $k = 1$.

Os limites para a densidade de acordo com as normas EN 14214 e RANP 45/2014 são bem parecidos e o banco de dados utilizado para o ajuste dos modelos contém amostras bem distribuídas próximas aos limites, especialmente no limite superior de ambas as normas e no limite inferior da EN 14214. Isso pode ter levado aos resultados máximos de acurácia, especificidade e sensibilidade obtidos na classificação das amostras de validação externa pelos quatro métodos de classificação para a norma EN 14214.

Considerando os limites da RANP 45/2014 para a densidade, os métodos SVM, KNN e TREE tiveram resultados máximos de acurácia, sensibilidade e especificidade, mas os resultados obtidos por ANN também foram satisfatórios.

Para a classificação da densidade tanto pelos limites da EN 14214 quando pela RANP 45/2014, recomenda-se o uso de TREE com base nos critérios secundários adotados neste trabalho. O modelo de TREE otimizado teve novamente Gini's Diversity Index como critério de decisão e o algoritmo Exact.

4.3 Resultados dos modelos preditivos

Os métodos de classificação com os melhores desempenhos no presente trabalho foram comparados com equações da literatura, conforme mostrado na Fig. 4.4. Visto que este é o primeiro trabalho que otimiza e aplica métodos de classificação direta em parâmetros de qualidade de biodiesel, é importante comparar o desempenho com métodos que já estão bem estabelecidos na literatura: os métodos de classificação indireta.

No presente trabalho, entende-se por classificação indireta a aplicação de um método de regressão para prever o valor da propriedade e realizar a classificação da amostra, com base nos limites de cada classe (especificações das normas adotadas). Tanto para os classificadores diretos (TREE e KNN) quando para os classificadores indiretos (equações empíricas da literatura) foram utilizados os FAMEs como entrada e os parâmetros de qualidade como saída. A comparação foi realizada apenas em relação às amostras de validação externa.

A classificação da viscosidade pela norma EN 14214 foi a única cujo classificador direto teve performance inferior às equações empíricas. Ainda que KNN tenha tido máxima acurácia e boas sensibilidade e especificidade, os três parâmetros foram máximos para os modelos 1 e 2.

Para as demais normas e propriedades, os classificadores diretos foram iguais ou melhores do que as equações empíricas. Para a classificação da viscosidade pela RANP 45/2014, o desempenho de TREE foi disparadamente melhor que os classificadores indiretos. O desempenho de KNN para classificação de estabilidade oxidativa pela ASTM D6751 também foi bem melhor que os métodos para classificação indireta.

Apenas três equações obtiveram desempenhos máximos na classificação indireta, que foram os modelos 1, 2 e 7. Alguns modelos tiveram especificidade zero e baixa acurácia, como os modelos 5, 6 para ASTM D6751 e o modelo 8 para EN 14214 e RANP 45/2014. Isso se deve às elevadas raízes dos erros médios quadráticos (RMSE) para as amostras de validação externa. Os modelos 5 e 6 tiveram RMSE de 2,15 h e 1,80 h respectivamente, para as amostras da estabilidade oxidativa pela ASTM D6751, enquanto que o modelo 8 teve RMSE 30,93 kg/m³ e 32,38 kg/m³ para as amostras de densidade das normas EN 14214 e RANP 45/2014.

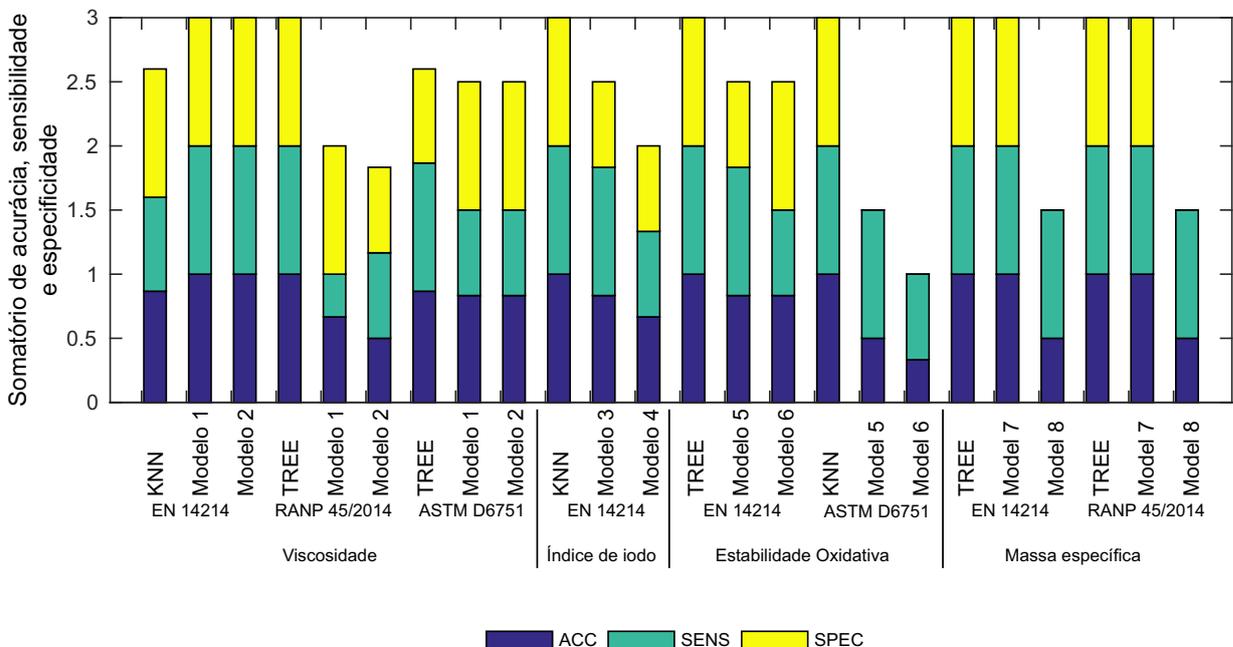


Figura 4.4: Comparação entre métodos de classificação direta (presente trabalho) e classificação indireta (equações da literatura) aplicada ao conjunto de validação externa para a classificação de biodiesel de acordo com as propriedades de viscosidade, índice de iodo, estabilidade oxidativa e massa específica

Como as amostras selecionadas para a validação externa tinham valores próximos aos limites de cada propriedade de acordo com cada norma, um alto RMSE pode ser comprometedor para a aplicação dessas equações para classificação indireta, o que ressalta a importância da aplicação dos classificadores diretos, desde que tenham sido treinados com um banco de dados com boa variabilidade dos dados.

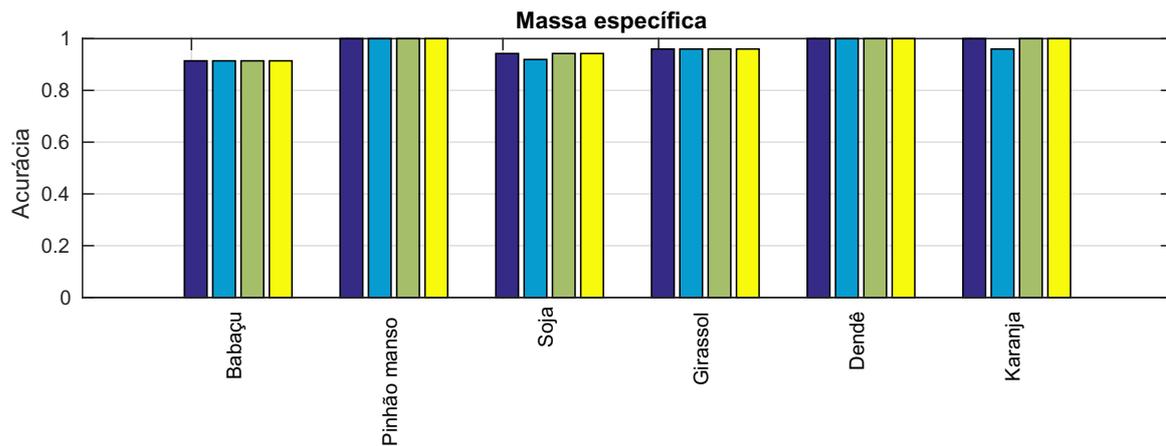
Analisando por matéria-prima, para a massa específica todos os algoritmos obtiveram resultado semelhante, com KNN apresentando resultado pouco inferior para soja e karanja (Figura 4.5a). Para a viscosidade, SVM e TREE foram os que obtiveram resultado máximo para a maioria das amostras (Figura 4.5b). ANN obteve valor superior aos outros algoritmos para algumas amostras na propriedade estabilidade oxidativa: babaçu, pinhão manso, soja, dendê, mistura com 60% de soja e para karanja. Para o índice de iodo ANN obteve melhor resultado para soja e dendê.

4.4 Software

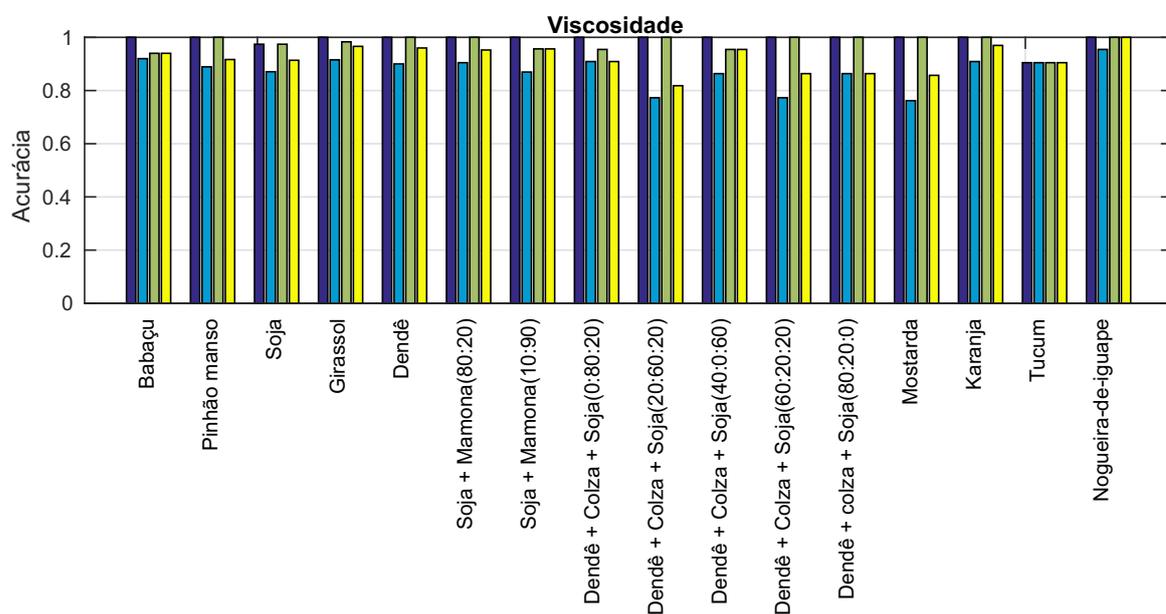
Foi desenvolvido um programa de computador de fácil utilização baseado na metodologia adotada, escolhendo os classificadores que obtiveram os melhores resultados. O programa está disponível na versão beta em: <<https://drive.google.com/drive/folders/0B6Ede4woZiikeFNRdlRncEJPQUk?usp=sharing>>.

É ainda necessário instalar o conjunto de bibliotecas do MATLAB necessário para executáveis (MATLAB Runtime R2016b (9.1)) de acordo com o sistema operacional usado, disponível em: <<https://www.mathworks.com/products/compiler/mcr.html>>

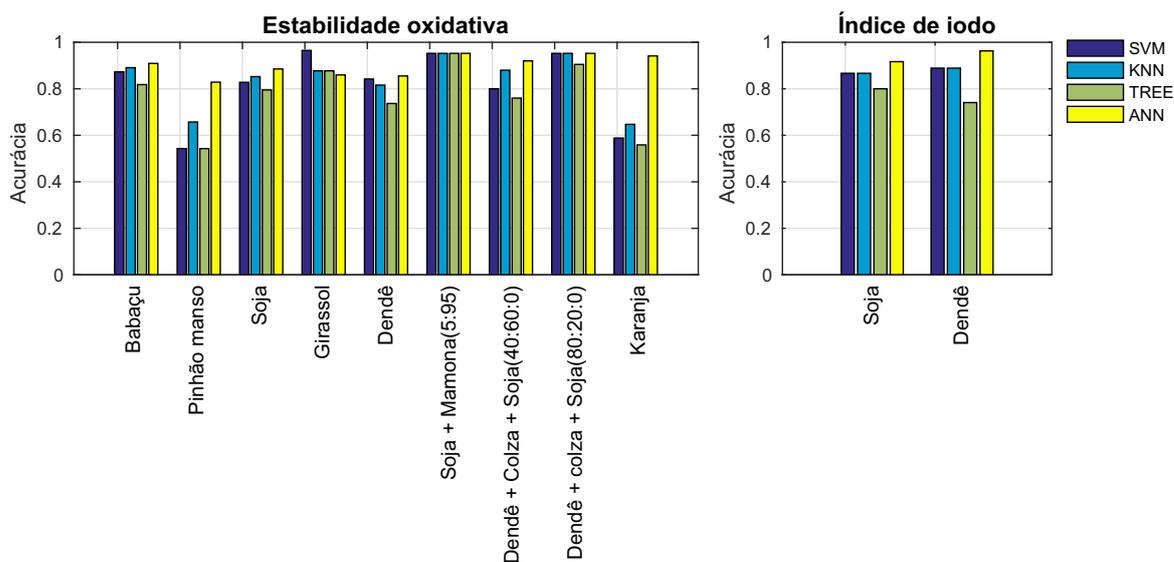
Pode-se escolher entre cada uma das especificações (ASTM D6751, EN 14214 ou RANP 45/2014) e verificar os resultados para as propriedades compreendidas. O usuário deve utilizar como entrada as porcentagens dos FAMES das amostras. Caso a amostra não esteja em conformidade pode-se entrar com outras amostras e escolher para o cálculo uma mistura binária ou ternária.



(a)



(b)



(c)

(d)

Figura 4.5: Resultados de classificação para as propriedades de massa específica (a), viscosidade (b), estabilidade oxidativa (c) e índice de iodo (d) do conjunto de teste de acordo com o tipo de matéria-prima utilizado.

Capítulo 5

Conclusão, Publicações e Trabalhos Futuros

5.1 Conclusão

Métodos de classificação direta foram otimizados e comparados no presente trabalho para classificação de viscosidade, índice de iodo, estabilidade oxidativa e densidade de amostras de biodiesel, de acordo com as normas EN 14214, RANP 45/2014 e ASTM D6751. Tomando por critério os melhores desempenhos e menor complexidade, os métodos TREE e KNN se revezaram com os melhores resultados para as classificações, garantindo máximas acurácia, sensibilidade e especificidade para as amostras de validação externa de quase todas as propriedades e normas.

Uma comparação entre os classificadores diretos e equações empíricas da literatura revelou que a performance dos classificadores indiretos pode ficar comprometida por altos valores de RMSE, especialmente quando as amostras de biodiesel têm valores das propriedades bem próximos aos limites das especificações consideradas, atribuindo a importância dos métodos de classificação direta, tais como os recomendados no presente trabalho, para as aplicações do problema abordado.

Finalmente um programa de computador foi desenvolvido baseado no método proposto para facilitar a verificação de amostras de óleo e forma rápida e com precisão a partir dos FAMES nas plantas de produção de biodiesel.

5.2 Trabalhos Publicados

Lopes, M. V.; Barradas Filho, A. O.; Barros, A.K.; Viegas, I.M.A.; Silva, L.C.O.; Marques, E.P.; Marques, A.L.B. Attesting compliance of biodiesel quality using composition data and classification methods. *Neural Computing and Applications*, 2017.

5.3 Trabalhos Futuros

Como os algoritmos de classificação mostraram-se eficientes para a verificação da conformidade do biodiesel de acordo com as normas, o próximo passo é apresentar formas de sugestão de misturas com diferentes matérias-primas que atendam os parâmetros e sejam viáveis economicamente para a região.

Referências Bibliográficas

AGARWAL, M.; SINGH, K.; CHAURASIA, S. P. Prediction of biodiesel properties from fatty acid composition using linear regression and ANN techniques. *Indian Chemical Engineer*, v. 52, n. 4, p. 347 – 361, 2010. 56

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS. *Resolução ANP no 45, DE 25.8.2014 - DOU 26.8.2014*. [S.l.], 2014. Acesso em: 13 Fevereiro 2015. Disponível em: <<http://www.anp.gov.br>>. 16, 30

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS. *Boletim Mensal do Biodiesel*. [S.l.], 2017. Acesso em: 04 julho 2016. Disponível em: <<http://www.anp.gov.br>>. 25, 26, 34, 35

AHMAD, M. et al. *Practical Handbook on Biodiesel Production and Properties*. [S.l.]: CRC Press, 2012. 24, 25, 33

AL-HAMAMRE, Z.; AL-SALAYMEH, A. Physical properties of (jojoba oil+biodiesel), (jojoba oil+diesel) and (biodiesel+diesel) blends. *Fuel*, v. 123, p. 175 – 188, 2014. 18, 53

ALVES, J. C. L.; HENRIQUES, C. B.; POPPI, R. J. Determination of diesel quality parameters using support vector regression and near infrared spectroscopy for an in-line blending optimizer system. *Fuel*, v. 97, p. 710 – 717, 2012. 20, 35

ALVES, J. C. L.; POPPI, R. J. Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). *Talanta*, v. 104, p. 155 – 161, 2013. 20, 35

ARISHOLM, E.; BRIAND, L. C.; JOHANNESSEN, E. B. A systematic and comprehensive investigation of methods to build and evaluate fault prediction models. *Journal of Systems and Software*, v. 83, n. 1, p. 2 – 17, 2010. SI: Top Scholars. 43

ARUMUGAM, S.; SRIRAM, G.; SUBRAMANIAN, P. S. Application of artificial intelligence to predict the performance and exhaust emissions of diesel engine using rapeseed oil methyl ester. *Procedia Engineering*, v. 38, p. 853 – 860, 2012. International Conference on Modeling Optimization and Computing. 53

ATABANI, A. et al. A comparative evaluation of physical and chemical properties of biodiesel synthesized from edible and non-edible oils and study on the effect of biodiesel blending. *Energy*, v. 58, p. 296 – 304, 2013. 17, 18, 33

ATABANI, A. et al. Effect of croton megalocarpus, calophyllum inophyllum, moringa oleifera, palm and coconut biodiesel-diesel blending on their physico-chemical properties. *Industrial Crops and Products*, v. 60, p. 130 – 137, 2014. 19

- ATABANI, A. et al. A comprehensive review on biodiesel as an alternative energy resource and its characteristics. *Renewable and Sustainable Energy Reviews*, v. 16, n. 4, p. 2070 – 2093, 2012. 17
- AYETOR, G. K.; SUNNU, A.; PARBEY, J. Performance evaluation of biodiesel-biodiesel blends in a dedicated CIDI engine. *International Journal of Renewable Energy Research*, v. 5, n. 1, 2015. 17
- BALA, B. Studies on biodiesels from transformation of vegetable oils for diesel engines. *Energy Education Science and Technology, SILA SCIENCE*, v. 15, n. 1/2, p. 1, 2005. 25
- BALABIN, R. M.; LOMAKINA, E. I.; SAFIEVA, R. Z. Neural network (ANN) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. *Fuel*, v. 90, n. 5, p. 2007 – 2015, 2011. 19
- BALABIN, R. M.; SAFIEVA, R. Z. Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data. *Analytica Chimica Acta*, v. 689, n. 2, p. 190 – 197, 2011. 19, 54
- BALABIN, R. M.; SAFIEVA, R. Z.; LOMAKINA, E. I. Gasoline classification using near infrared (NIR) spectroscopy data: Comparison of multivariate techniques. *Analytica Chimica Acta*, v. 671, n. 1-2, p. 27–35, 2010. 54
- BALABIN, R. M.; SAFIEVA, R. Z.; LOMAKINA, E. I. Near-infrared (NIR) spectroscopy for motor oil classification: From discriminant analysis to support vector machines. *Microchemical Journal*, v. 98, n. 1, p. 121 – 128, 2011. 21
- BALAT, M. Potential alternatives to edible oils for biodiesel production - a review of current work. *Energy Conversion and Management*, v. 52, n. 2, p. 1479 – 1492, 2011. 47
- BANKOVIĆ-ILIĆ, I. B.; STAMENKOVIĆ, O. S.; VELJKOVIĆ, V. B. Biodiesel production from non-edible plant oils. *Renewable and Sustainable Energy Reviews*, v. 16, n. 6, p. 3621 – 3647, 2012. 33
- BARABÁS, I.; TODORUȚ, I.-A. Biodiesel quality, standards and properties. In: _____. *Biodiesel- Quality, Emissions and By-Products*. [S.l.]: InTech, 2011. p. 3 – 28. 28, 29
- BARNWAL, B.; SHARMA, M. Prospects of biodiesel production from vegetable oils in india. *Renewable and Sustainable Energy Reviews*, v. 9, n. 4, p. 363 – 378, 2005. 33
- BARRADAS FILHO, A. O. *Avaliação de matérias-primas para qualidade de biodieseis pela predição de propriedades físico-químicas*. Tese (Doutorado) — Universidade Federal do Maranhão, 2015. 47, 54, 56
- BARRADAS FILHO, A. O. et al. Application of artificial neural networks to predict viscosity, iodine value and induction period of biodiesel focused on the study of oxidative stability. *Fuel*, v. 145, p. 127 – 135, 2015. 19, 47, 48, 53, 56, 60
- BENJUMEA, P.; AGUDELO, J.; AGUDELO, A. Basic properties of palm oil biodiesel-diesel blends. *Fuel*, v. 87, n. 10-11, p. 2069 – 2075, 2008. 19
- BISHOP, C. M. *Pattern recognition and machine learning*. [S.l.]: springer, 2006. 35

BRASIL. *Lei nº 11.097, de 13 de janeiro de 2005*: Dispõe sobre a introdução do biodiesel na matriz energética brasileira; altera as leis nºs 9.478, de 6 de agosto de 1997, 9.847, de 26 de outubro de 1999 e 10.636, de 30 de dezembro de 2002; e dá outras providências. Brasília, DF, 2005. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/lei/111097.htm>. 16

BRASIL. *Lei nº 13.263, de 23 de março de 2016*: Altera a lei nº 13.033, de 24 de setembro de 2014, para dispor sobre os percentuais de adição de biodiesel ao óleo diesel comercializado no território nacional. Brasília, DF, 2016. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/lei/L13263.htm>. 16

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. *Usos de biodiesel no Brasil e no mundo*. 1. ed. [S.l.], 2015. 30, 31, 32

BRITISH PETROLEUM. *BP statistical review of world energy June 2017*. [S.l.]: London: British Petroleum, 2017. 17

BRUDZEWSKI, K. et al. Gasoline quality prediction using gas chromatography and FTIR spectroscopy: An artificial intelligence approach. *Fuel*, v. 85, n. 4, p. 553 – 558, 2006. 35

CANDEIA, R. et al. Influence of soybean biodiesel content on basic properties of biodiesel-diesel blends. *Fuel*, v. 88, n. 4, p. 738 – 743, 2009. 19

CANDEIA, R. A. *Biodiesel de soja: síntese, degradação e misturas binárias*. Tese (Doutorado) — Universidade Federal da Paraíba, 2008. 47

CHANG, C.-C.; LIN, C.-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, v. 2, p. 27:1–27:27, 2011. Software available at <<http://www.csie.ntu.edu.tw/~cjlin/libsvm>>. 52

CHAVARRIA-HERNANDEZ, J. C.; PACHECO-CATALÁN, D. E. Predicting the kinematic viscosity of FAMES and biodiesel: Empirical models. *Fuel*, v. 124, p. 212 – 220, 2014. 20, 27

CHAWLA, N. V. et al. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357, 2002. 44, 47

CORNELL, J. A. *Experiments with mixtures: designs, models, and the analysis of mixture data*. [S.l.]: Wiley, 2002. (Wiley series in probability and mathematical statistics. Probability and mathematical statistics). 44, 45

COSTA, A. et al. Sustainability analysis of biodiesel production: A review on different resources in Brazil. *Renewable and Sustainable Energy Reviews*, v. 27, p. 407 – 412, 2013. 34

COSTA, A. C. A. da; JUNIOR, N. P.; ARANDA, D. A. G. The situation of biofuels in Brazil: New generation technologies. *Renewable and Sustainable Energy Reviews*, v. 14, n. 9, p. 3041 – 3049, 2010. 25

COVER, T.; HART, P. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, v. 13, n. 1, p. 21–27, January 1967. 40

- CREMONEZ, P. A. et al. Biodiesel production in Brazil: Current scenario and perspectives. *Renewable and Sustainable Energy Reviews*, v. 42, p. 415 – 428, 2015. 33, 34
- DAHIYA, A. *Bioenergy: biomass to biofuels*. [S.l.]: Academic Press, 2014. 26
- DAL BELO LEITE, J. G. et al. Socioeconomic and environmental assessment of biodiesel crops on family farming systems in Brazil. *Agricultural Systems*, v. 133, p. 22 – 34, 2015. 33
- DALL'AGNOL, A. *Por que fazemos biodiesel de soja*. [S.l.], 2015. Acesso em: 06 julho 2015. Disponível em: <<http://www.biodieselbr.com/noticias/columnistas/convidado/por-que-fazemos-biodiesel-de-soja.htm>>. 25
- DANTAS, M. et al. Evaluation of the oxidative stability of corn biodiesel. *Fuel*, v. 90, n. 2, p. 773 – 778, 2011. 19
- DANTAS, M. et al. Biodiesel from soybean oil, castor oil and their blends. *Journal of Thermal Analysis and Calorimetry*, Springer Netherlands, v. 106, n. 2, p. 607–611, 2011. 19, 47
- DANTAS, M. B. *Blendas de biodiesel: propriedades de fluxo, estabilidade térmica e oxidativa e monitoramento durante armazenamento*. Tese (Doutorado) — Universidade Federal da Paraíba, 2010. 47
- DEMIRBAS, A. Biodiesel production from vegetable oils via catalytic and non-catalytic supercritical methanol transesterification methods. *Progress in Energy and Combustion Science*, v. 31, n. 5-6, p. 466 – 487, 2005. ISSN 0360-1285. 33
- DEMIRBAS, A. *Biodiesel a Realistic Fuel Alternative for Diesel Engines*. [S.l.]: Springer, 2008. 24, 25, 33, 47
- DOMINGOS, P. Metacost: A general method for making classifiers cost-sensitive. In: *In Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining*. [S.l.]: ACM Press, 1999. p. 155 – 164. 43, 44
- EZAWA, K.; SINGH, M.; NORTON, S. W. Learning goal oriented bayesian networks for telecommunications risk management. In: *In Proceedings of the 13th International Conference on Machine Learning*. [S.l.]: Morgan Kaufmann, 1996. p. 139 – 147. 43
- FILGUEIRAS, P. R. et al. Determination of API gravity, kinematic viscosity and water content in petroleum by ATR-FTIR spectroscopy and multivariate calibration. *Fuel*, v. 116, p. 123 – 130, 2014. 35
- GUI, M.; LEE, K.; BHATIA, S. Feasibility of edible oil vs. non-edible oil vs. waste edible oil as biodiesel feedstock. *Energy*, v. 33, n. 11, p. 1646 – 1653, 2008. 33
- HALL, T. et al. A systematic literature review on fault prediction performance in software engineering. *Software Engineering, IEEE Transactions on*, v. 38, n. 6, p. 1276–1304, Nov 2012. 43
- HAYKIN, S.; LIPPMANN, R. *Neural Networks, A Comprehensive Foundation*. 2. ed. [S.l.]: Prentice Hall, 1999. 38

- HOEKMAN, S. K. et al. Review of biodiesel composition, properties, and specifications. *Renewable and Sustainable Energy Reviews*, v. 16, n. 1, p. 143 – 169, 2012. 16, 27, 47
- HONG, I. K.; JEON, G. S.; LEE, S. B. Prediction of biodiesel fuel properties from fatty acid alkyl ester. *Journal of Industrial and Engineering Chemistry*, v. 20, n. 4, p. 2348 – 2353, 2014. 20, 27, 28, 56
- HONG, I. K.; LEE, J. R.; LEE, S. B. Fuel properties of canola oil and lard biodiesel blends: Higher heating value, oxidative stability, and kinematic viscosity. *Journal of Industrial and Engineering Chemistry*, v. 22, p. 335 – 340, 2015. 18
- HSU, C. wei; CHANG, C. chung; LIN, C. jen. *A practical guide to support vector classification*. 2010. Disponível em: <<http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>>. 50, 52
- JAPKOWICZ, N. The class imbalance problem: Significance and strategies. In: *In Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI*. [S.l.: s.n.], 2000. p. 111 – 117. 43, 44
- KEGL, B.; KEGL, M.; PEHAN, S. *Green Diesel Engines: Biodiesel Usage in Diesel Engines*. [S.l.]: Springer, 2013. 27, 60
- KHANMOHAMMADI, M. et al. Quality based classification of gasoline samples by ATR-FTIR spectrometry using spectral feature selection with quadratic discriminant analysis. *Fuel*, v. 111, p. 96 – 102, 2013. 21
- KIVEVELE, T. T.; MBARAWA, M. M. Experimental investigations of oxidation stability of biodiesel produced from manketti seeds oil (*schinziophyton rautanenii*). *Energy & Fuels*, v. 25, n. 5, p. 2341–2346, 2011. 47
- KNOTHE, G.; STEIDLEY, K. R. Kinematic viscosity of biodiesel fuel components and related compounds. influence of compound structure and comparison to petrodiesel fuel components. *Fuel*, v. 84, n. 9, p. 1059 – 1065, 2005. 24, 60
- KUBAT, M.; HOLTE, R. C.; MATWIN, S. Machine learning for the detection of oil spills in satellite radar images. *Machine Learning*, v. 30, n. 2, p. 195–215, Feb 1998. 43
- LEWIS, D. D.; CATLETT, J. Heterogeneous uncertainty sampling for supervised learning. In: *In Proceedings of the Eleventh International Conference on Machine Learning*. [S.l.]: Morgan Kaufmann, 1994. p. 148 – 156. 43
- LING, C. et al. Data mining for direct marketing: Problems and solutions. In: *In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98*. [S.l.]: AAAI Press, 1998. p. 73 – 79. 44
- LÔBO, I. P.; FERREIRA, S. L. C.; CRUZ, R. S. d. Biodiesel: parâmetros de qualidade e métodos analíticos. *Química Nova*, v. 32, p. 1596 – 1608, 2009. 25, 29, 53
- MARTÍN, C. et al. Fractional characterisation of jatropha, neem, moringa, trisperma, castor and candlenut seeds as potential feedstocks for biodiesel production in cuba. *Biomass and Bioenergy*, v. 34, n. 4, p. 533 – 538, 2010. 47

- MEHER, L.; SAGAR, D. V.; NAIK, S. Technical aspects of biodiesel production by transesterification - a review. *Renewable and Sustainable Energy Reviews*, v. 10, n. 3, p. 248 – 268, 2006. 29
- MENG, X.; JIA, M.; WANG, T. Neural network prediction of biodiesel kinematic viscosity at 313 K. *Fuel*, v. 121, p. 133–140, 2014. 20, 28, 47
- MOFIJUR, M. et al. Comparative evaluation of performance and emission characteristics of moringa oleifera and palm oil based biodiesel in a diesel engine. *Industrial Crops and Products*, v. 53, p. 78 – 84, 2014. 19
- MOSER, B. R. Influence of blending canola, palm, soybean, and sunflower oil methyl esters on fuel properties of biodiesel. *Energy and Fuels*, v. 22, n. 6, p. 4301–4306, 2008. 18
- MOSER, B. R. Influence of extended storage on fuel properties of methyl esters prepared from canola, palm, soybean and sunflower oils. *Renewable Energy*, v. 36, n. 4, p. 1221 – 1226, 2011. 47
- NATIONAL RENEWABLE ENERGY LABORATORY. *Biodiesel Handling and Use Guide: Fourth Edition*. 4. ed. [S.l.]: National Renewable Energy Laboratory, 2008. 26, 29, 30
- NOGUEIRA, C. A. et al. Densities and viscosities of binary mixtures of babassu biodiesel + cotton seed or soybean biodiesel at different temperatures. *Journal of Chemical & Engineering Data*, v. 55, n. 11, p. 5305–5310, 2010. 47
- OGHENEJOBOR, K.; UMUKORO, P. Comparative analysis of fuel characteristics of biodiesel produced from selected oil-bearing seeds in nigeria. *European Journal of Scientific Research*, v. 58, n. 2, p. 238–246, 2011. Cited By 7. 19
- ÖZÇELİK, A. E.; AYDOĞAN, H.; ACAROĞLU, M. Determining the performance, emission and combustion properties of camelina biodiesel blends. *Energy Conversion and Management*, v. 96, p. 47 – 57, 2015. 19
- PADHI, S. K. *Preparation and Characterization of Biodiesel from Non-Edible Oils*. Tese (Doutorado) — National Institute of Technology, 2010. 47
- PARK, J.-Y. et al. Blending effects of biodiesels on oxidation stability and low temperature flow properties. *Bioresource Technology*, v. 99, n. 5, p. 1196 – 1203, 2008. 47, 56
- PONTES, M. J. C. et al. Screening analysis to detect adulteration in diesel/biodiesel blends using near infrared spectrometry and multivariate classification. *Talanta*, v. 85, n. 4, p. 2159 – 2165, 2011. 21
- PROVOST, F.; FAWCETT, T. Robust classification for imprecise environments. *Machine Learning*, v. 42, n. 3, p. 203–231, Mar 2001. 43
- QI, D.; LEE, C. Influence of soybean biodiesel content on basic properties of biodiesel-diesel blends. *Journal of the Taiwan Institute of Chemical Engineers*, v. 45, n. 2, p. 504 – 507, 2014. 19

- RAMIREZ-VERDUZCO, L. F. Density and viscosity of biodiesel as a function of temperature: Empirical models. *Renewable and Sustainable Energy Reviews*, v. 19, p. 652–665, 2013. 56
- RAMÍREZ-VERDUZCO, L. F.; RODRÍGUEZ-RODRÍGUEZ, J. E.; JARAMILLO-JACOB, A. del R. Predicting cetane number, kinematic viscosity, density and higher heating value of biodiesel from its fatty acid methyl ester composition. *Fuel*, v. 91, n. 1, p. 102 – 111, 2012. 20, 27, 28
- ROJAS-GONZÁLEZ, A.; GIRÓN-GALLEGO, E. Operation variables in the animal-fat transesterification process: a review. *Ingeniería y Universidad*, scieloco, v. 15, p. 197 – 218, 2011. 33
- ROMANO, S. D.; SORICHETTI, P. A. *Dielectric spectroscopy in biodiesel production and characterization*. [S.l.]: Springer Science & Business Media, 2010. 28, 29, 32
- SANJID, A. et al. Production of palm and jatropha based biodiesel and investigation of palm-jatropha combined blend properties, performance, exhaust emission and noise in an unmodified diesel engine. *Journal of Cleaner Production*, v. 65, p. 295 – 303, 2014. 17
- SANTOS, A. G. et al. Characterization and kinetic study of sunflower oil and biodiesel. *Journal of Thermal Analysis and Calorimetry*, Springer Netherlands, v. 106, n. 3, p. 747–751, 2011. ISSN 1388-6150. 19
- SANTOS, J. R. d. J. *Biodiesel de babaçu: avaliação térmica, oxidativa e misturas binárias*. Tese (Doutorado) — Universidade Federal da Paraíba, 2008. 47
- SANTOS, N. et al. Rheology and MT-DSC studies of the flow properties of ethyl and methyl babassu biodiesel and blends. *Journal of Thermal Analysis and Calorimetry*, Springer Netherlands, v. 106, n. 2, p. 501–506, 2011. 19
- SARIN, A. *Biodiesel: production and properties*. [S.l.]: Royal Society of Chemistry, 2012. 16, 17, 27, 28, 60
- SARIN, R. et al. Jatropha-palm biodiesel blends: An optimum mix for asia. *Fuel*, v. 86, n. 10, p. 1365 – 1371, 2007. 19, 47
- SAXENA, P.; JAWALE, S.; JOSHIPURA, M. H. A review on prediction of properties of biodiesel and blends of biodiesel. *Procedia Engineering*, v. 51, p. 395 – 402, 2013. Chemical, Civil and Mechanical Engineering Tracks of 3rd Nirma University International Conference on Engineering (NUiCONE2012). 27, 28
- SCHÖLKOPF, B.; SMOLA, A. J. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. [S.l.]: MIT press, 2002. 36, 37, 39
- SHI, F. et al. Optimization of biodiesel injection parameters based on support vector machine. *Mathematical Problems in Engineering*, 2013. 20, 35
- SILITONGA, A. et al. Overview properties of biodiesel diesel blends from edible and non-edible feedstock. *Renewable and Sustainable Energy Reviews*, v. 22, p. 346 – 360, 2013. 19

-
- SILVA, R. B. da et al. Catalysts of Cu(II) and Co(II) ions adsorbed in chitosan used in transesterification of soy bean and babassu oils - a new route for biodiesel syntheses. *Bioresource Technology*, v. 99, n. 15, p. 6793 – 6798, 2008. 47
- STOYTICHEVA, M. et al. *Biodiesel: feedstocks and processing technologies*. [S.l.]: Intech, 2011. 16, 17
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introduction to Data Mining*. [S.l.]: Addison Wesley, 2006. 35, 41, 42
- TAVARES, M. et al. Sunflower biodiesel. *Journal of Thermal Analysis and Calorimetry*, Springer Netherlands, v. 106, n. 2, p. 575–579, 2011. 19
- VALENTE, O. S. et al. Physical-chemical properties of waste cooking oil biodiesel and castor oil biodiesel blends. *Fuel*, v. 90, n. 4, p. 1700 – 1702, 2011. 19
- VAPNIK, V. *The Nature of Statistical Learning Theory*. [S.l.]: Springer-Verlag, 1995. 35
- VERAS, G. et al. Classification of biodiesel using NIR spectrometry and multivariate techniques. *Talanta*, v. 83, n. 2, p. 565 – 568, 2010. 19
- WANG, L. bing et al. Influence of fatty acid composition of woody biodiesel plants on the fuel properties. *Journal of Fuel Chemistry and Technology*, v. 40, n. 4, p. 397 – 404, 2012. 56
- WEN, Z.; JOHNSON, M. B. Microalgae as a feedstock for biofuel production. Virginia Cooperative Extension, 2009. 33
- YANG, H. et al. Neural network prediction of cetane number and density of diesel fuel from its chemical composition determined by LC and GC-MS. *Fuel*, v. 81, n. 1, p. 65 – 74, 2002. 20, 28, 53
- YUSUP, S.; KHAN, M. Basic properties of crude rubber seed oil and crude palm oil blend as a potential feedstock for biodiesel production with enhanced cold flow characteristics. *Biomass and Bioenergy*, v. 34, n. 10, p. 1523 – 1526, 2010. 18, 33

Apêndice A

Software

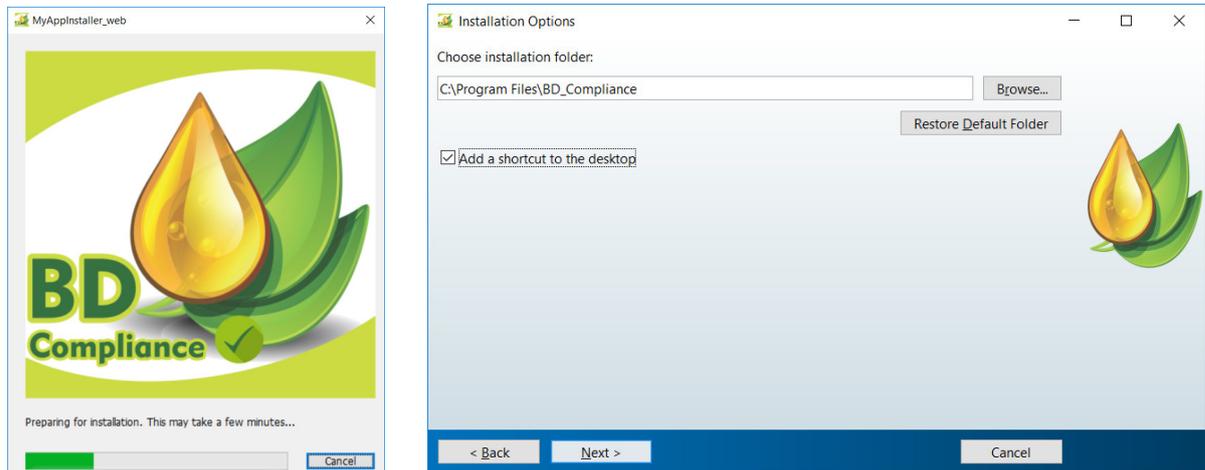


Figura A.1: Telas de instalação.

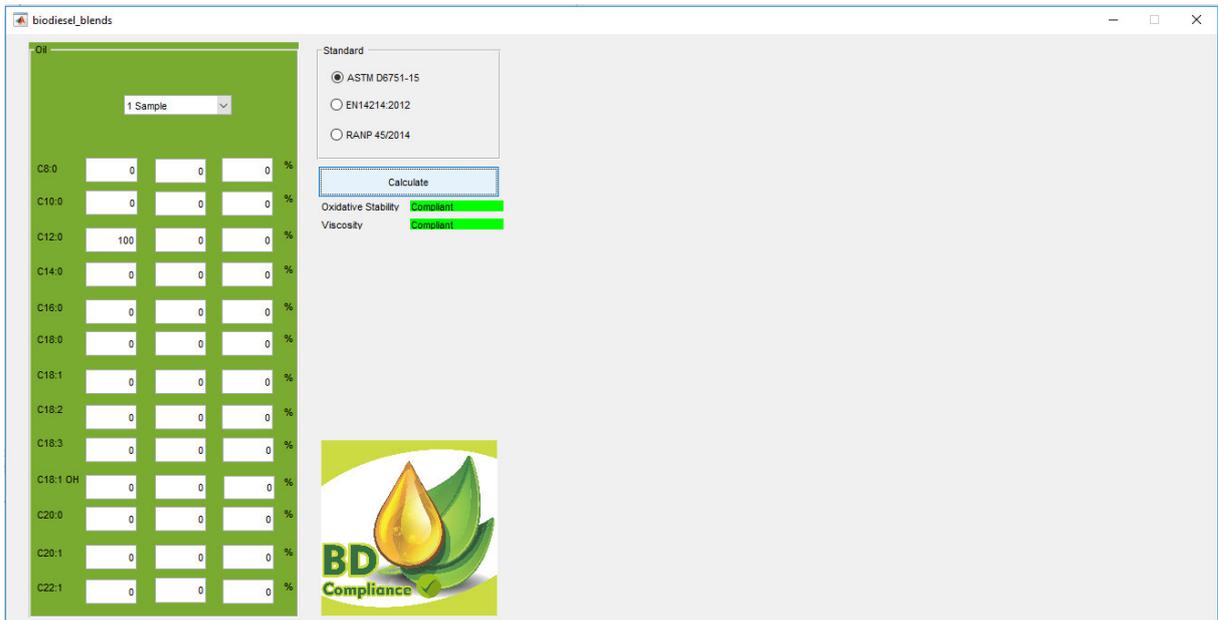


Figura A.2: Classificação para uma amostra.

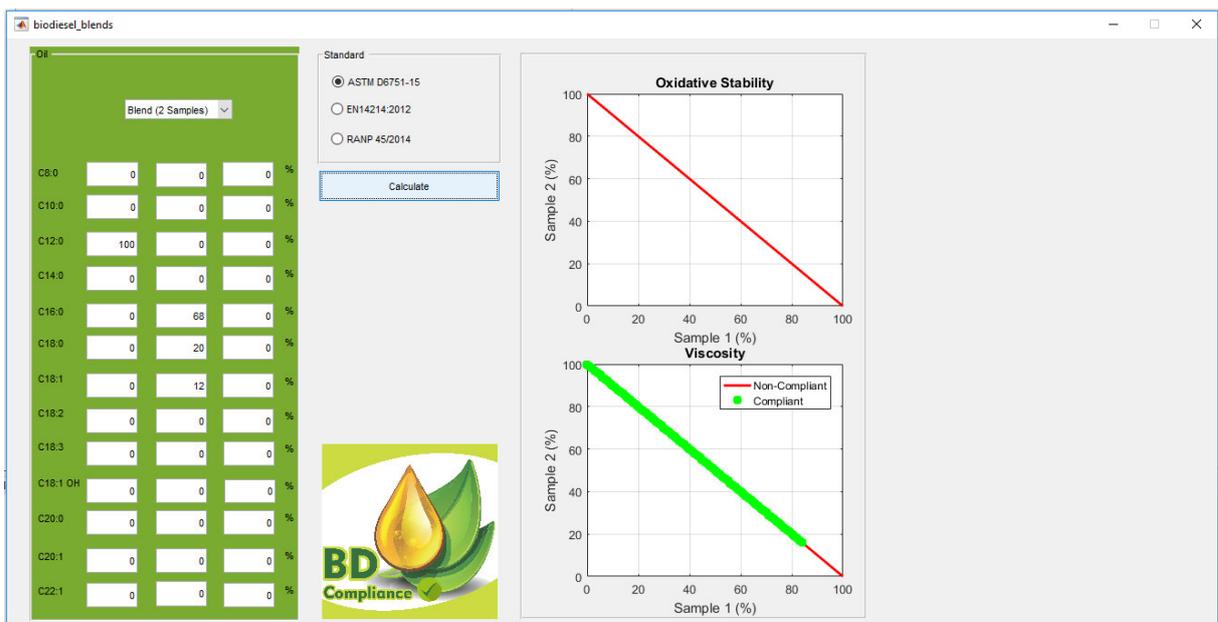


Figura A.3: Classificação para duas amostras (mistura binária).

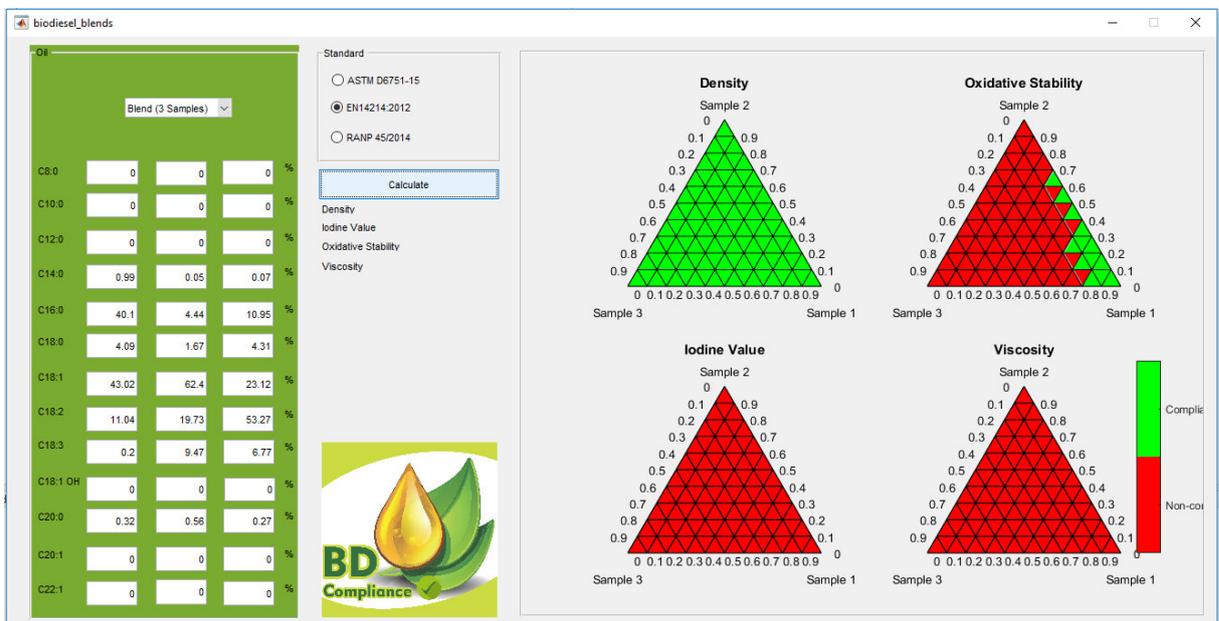


Figura A.4: Classificação para três amostras (mistura ternária).