Revista Brasileira de Informática na Educação – RBIE
Brazilian Journal of Computers in Education
(ISSN online: 2317-6121; print: 1414-5685)
https://sol.sbc.org.br/journals/index.php/rbie

Submission: 18/Aug/2025; 1st round notif.: 18/Aug/2025; New version: 18/Aug/2025; 2nd round notif.: 20/Oct/2025;
Camera ready: 20/Oct/2025; Edition review: 20/Oct/2025; Available online: 17/Nov/2025; Published: 17/Nov/2025;

Identificação de tecnologias emergentes de etanol a partir de dados de patentes brasileiras usando ML

Title: Identification of emerging ethanol technologies from Brazilian patent data using ML

Osvaldo Carvalho dos Santos Neto
Escola de Artes, Ciências e Humanidades - USP
ORCID: 0000-0000-0000-0000
osvaldocsantos@usp.br

Gabriel da Silva Simões
Escola de Artes, Ciências e Humanidades - USP
ORCID: 0000-0000-0000-0000
gabriel.s.simoeslp@usp.br

Luan Pereira Pinheiro
Escola de Artes, Ciências e Humanidades - USP
NUSP: 13672471
luanpinheiro@usp.br

Luis Felipe Pinheiro Felisberto
Escola de Artes, Ciências e Humanidades - USP
ORCID: 0000-0000-0000-0000
luis.felipe@usp.br

Resumo

Com o grande enfoque dado aos Objetivos de Desenvolvimento Sustentável (ODS), o estudo e uso de tecnologias sustentáveis se torna imprescindível. Um grande aliado nesse desenvolvimento é o etanol, que é um combustível sustentável produzido em larga escala no Brasil. Sabe-se, hoje, que o País tem capacidade para aperfeiçoar as tecnologias em uso e atingir grandes ganhos na produção de etanol, com o advento de algumas tecnologias em desenvolvimento, sem a necessidade de expansão das áreas cultivadas com cana-de-açúcar (Vian, 2022). A partir da análise de patentes brasileiras, faremos uma análise de tecnologias emergentes de etanol utilizando técnicas sofisticadas de aprendizado de máquina semissupervisionado, com aplicação dos algoritmos random forest e support vector machine.

Palavras-chave: *Etanol; Patente; Aprendizado de Máquina; Random Forest; Support Vector Machine; Emergente.*

Abstract

With the strong focus on the Sustainable Development Goals (SDGs), the study and use of sustainable technologies has become essential. A major ally in this development is ethanol, a sustainable fuel produced on a large scale in Brazil. It is now known that the country has the capacity to improve the technologies in use and achieve significant gains in ethanol production with the advent of some technologies under development, without the need to expand the areas cultivated with sugarcane (Vian, 2022). Based on an analysis of Brazilian patents, we will analyze emerging ethanol technologies using sophisticated semi-supervised machine learning techniques, applying random forest and support vector machine algorithms.

Keywords: *Ethanol; Patent; Machine Learning; Random Forest; Support Vector Machine; Emerging; Brazil.*

Cite as: Neto, O. C. S., Simões, G. S., Pinheiro, L. P. & Felisberto, L. F. P. (2025). Identificação de tecnologias emergentes de etanol a partir de dados de patentes usando ML. Revista Brasileira de Informática na Educação.

1 Introdução

O etanol ou álcool etílico ou apenas álcool é uma substância de grande importância na indústria em geral. No setor energético ele se destaca como uma alternativa menos poluente e renovável comparada aos combustíveis fósseis já que, dentre outras vantagens, emite 73% a menos de CO2 que a gasolina (Ferreira, 2009). Segundo a Barros (2021), a indústria alcoolquímica que utiliza o etanol como matéria-prima para a fabricação de produtos poderá vir a substituir a petroquímica, colocando o álcool etílico como uma opção de matéria prima acima do petróleo.

O Brasil é o maior produtor do mundo de cana-de-açúcar e na safra 2020/2021 foi responsável pela produção de 654,5 milhões de toneladas, destinados à produção de 41,2 milhões de toneladas de açúcar e 29,7 bilhões de litros de etanol (CONAB - Companhia Nacional de Abastecimento, 2021) sendo o estado de São Paulo líder na produção no país com 425,6 milhões de toneladas colhidas e 14,7 milhões m³ de etanol produzidos segundo o SEADE - Sistema Estadual de Análise de Dados (2021).

Esses dados colocam o etanol como um produto de extrema importância para o avanço em relação aos Objetivos de Desenvolvimento Sustentável (ODS), em particular o ODS 7: energia limpa e sustentável (ONU - Organização das Nações Unidas, 2023). Sendo assim, o estudo de tecnologias de etanol emergentes desenvolvidas no Brasil se faz relevante no cenário brasileiro e global, além de servir de apoio para o avanço do ODS 9 (indústria, inovação e infraestrutura), especialmente nos tópicos 9.5 e 9.b.

2 Fundamentos Teóricos

Para o desenvolvimento do presente trabalho, alguns fundamentos teóricos são essenciais, tais como o conceito de patentes e conceitos relacionados aos algoritmos de aprendizado de máquina usados na identificação das tecnologias emergentes. Esses fundamentos serão apresentados nesta seção.

Patente é um título de propriedade temporária sobre uma invenção ou modelo de utilidade, outorgado pelo Estado aos inventores (SEBRAE - Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2017). Em 2023, o INPI - Instituto Nacional da Propriedade Industrial (2023) registrou o depósito de 27.918 patentes, um crescimento de 2,9% em relação ao período anterior.

Uma predição é o resultado de uma análise que permite inferir previamente conclusões sobre o futuro, sendo essas inferências consideradas valiosas para a indústria para realizar a tomada de decisões de forma a minimizar os riscos e custos, e atingir os objetivos suavemente (LIN, 2021, p.74 citado em (Lee et al., 2022, p. 5)), neste trabalho, será comparado o desempenho de um algoritmo de random forest e de um support vector machine na realização de predições assertivas.

Uma árvore de decisão é um tipo de diagrama hierárquico que ajuda a visualizar etapas, decisões e o possível resultado de cada decisão popular em machine learning para tarefas de classificação e regressão de modelos (IBM - INTERNATIONAL BUSINESS MACHINES, 2023b). Random forest é um algoritmo utilizado para tarefas de classificação e regressão que combina a saída de múltiplas árvores de decisão para alcançar um único resultado (IBM - INTERNATIONAL

BUSINESS MACHINES, 2023a).

Uma máquina de vetores de suporte (SVM) é um algoritmo supervisionado de aprendizado de máquina que classifica dados encontrando uma linha ou hiperplano ótimo que maximiza a distância entre cada classe em um espaço N-dimensional (IBM - INTERNATIONAL BUSINESS MACHINES, 2023c).

A base da possibilidade de inferir quais são as tecnologias promissoras vem da implicação que a rede de patentes conectadas pelas citações se comporta de forma similar a um grafo direcionado evoluindo, cujas conexões representam referências a uma tecnologia anterior como base para criação de uma nova, logo os vértices de origem mais centrais podem ser destacados como fonte de inovação. Logo, algoritmos de aprendizado supervisionado como Random Forest e Support Vector Machine podem ser treinados a partir de dados anteriores para reconhecer e destacar as características desses vértices, de forma a obter as patentes, autores e tecnologias citadas.

Mais especificamente, neste trabalho faremos uso de algoritmos de aprendizado semissupervisionado, que consistem em algoritmos que fazem uso das técnicas de aprendizado supervisionado em um pequeno conjunto de dados, como base para um aprendizado não supervisionado para um conjunto grande de dados (Zhu, 2005). São usados principalmente quando adquirir rótulos para os dados é difícil (e.g. classificação de potencial de patentes a partir de revisão humana).

3 Trabalhos Relacionados

No estudo de Chung et al. (2020), "Early detection of valuable patents using a deep learning model: Case of semiconductor industry", é proposto um modelo de aprendizado profundo combinando CNN e LSTM para extrair características semânticas de patentes, classificando-as em três níveis de valor com base em citações futuras anuais. O modelo apresentou mais de 75% de precisão na identificação de patentes promissoras no setor de semicondutores.

De forma complementar, Kwon e Geum (2020) utilizaram 17 indicadores de patentes e técnicas de machine learning para prever invenções promissoras, destacando que a qualidade da acumulação de conhecimento é o preditor mais relevante para o sucesso das invenções (Kwon & Geum, 2020).

Além disso, o estudo de Park et al. (2021) avaliou patentes de circuitos integrados por meio de uma estratégia multidimensional de indicadores e diferentes modelos de machine learning, verificando que o algoritmo Random Forest alcançou precisão e acurácia superiores a 95% na classificação de patentes de alto valor (Park et al., 2021). No domínio de veículos elétricos, Li et al. (2021) abordaram a previsão de citações futuras como um problema de classificação, utilizando SVM otimizado para identificar patentes altamente citadas e mapear frentes tecnológicas emergentes (Li et al., 2021).

Em patentes de biomedicina têxtil, Zhao et al. (2021) desenvolveram o modelo BioTexVal, integrando BERT e múltiplos algoritmos de machine learning para prever o valor das patentes, alcançando aproximadamente 88% de acurácia ao treinar com 113.428 patentes (Zhao et al., 2021). No contexto brasileiro, Kazmi et al. (2022) investigaram o papel do país no desenvolvimento de tecnologias para produção de etanol de segunda geração por meio da análise de patentes publi-

cadas entre 2006 e 2015 (Kazmi et al., 2022).

No estudo "Forecasting emerging technologies: A supervised learning approach through patent analysis" de Kyebambe et al. (2017) desenvolveu um algoritmo para rotular automaticamente patentes como "emergentes" ou "não emergentes" e usar esses dados para treinar modelos de aprendizado de máquina supervisionado. No entanto, diferente do nosso estudo, o artigo de Kyebambe busca identificar ondas tecnológicas emergentes enquanto nós buscamos identificar uma tecnologia emergente apenas.

Por fim, o trabalho de Park et al. (2020) apresentou uma abordagem semi-supervisionada para identificar tecnologias emergentes, combinando um pequeno conjunto de patentes rotuladas por especialistas com um grande conjunto não rotulado, permitindo rotular automaticamente muitas patentes e facilitar a descoberta de inovações promissoras (Park et al., 2020). Nosso estudo se inspira nesta metodologia, buscando preencher a lacuna existente na análise de patentes recentes de tecnologias emergentes de etanol no Brasil.

4 Metodologia

O desenvolvimento deste estudo demandou, inicialmente, uma etapa de coleta e organização de dados. Para isso, foram utilizadas as bases de dados de patentes do Instituto Nacional da Propriedade Industrial (INPI) e da Organização Mundial da Propriedade Intelectual (WIPO).

Na base do INPI, empregou-se a palavra-chave "etanol", que apresentou resultados satisfatórios, embora limitações estruturais do sítio impeçam a formulação de consultas mais complexas. Para assegurar comparabilidade metodológica, selecionamos os mesmos códigos IPC utilizados por Perrone et al. (2011), cujo foco também recai sobre tecnologias associadas à produção de etanol. Dessa forma, incorporamos uma estratégia validada previamente na literatura. Adicionalmente, diversas consultas foram testadas na WIPO até se chegar à expressão de busca final, reproduzida abaixo:

```
IC:(C12P OR C12N OR C10L OR C07C OR A23B) AND
FP:(metanol OR methanol OR sugar OR etanol OR ethanol OR cana OR stover
    OR celulose OR bagasse OR madeira OR wood OR wooden OR cellulose
   OR bagaco OR beterraba OR beet OR sugarcane OR sucrose OR acucar*
   OR melaco OR melaco OR alcool* OR bioetanol OR bioethanol OR etil*
   OR milho OR corn OR soy OR soybean OR soja OR cereal OR trigo OR
   starch OR lignocellulose OR lignocelulose OR palha OR res?duo* OR
   biomass OR biomassa)
```

Ambas as bases retornam informações como título da patente, inventores, data e número de publicação. Este último funciona como identificador global, sendo utilizado para compor um arquivo CSV. Realizou-se, então, um tratamento desses identificadores, pois o formato utilizado pela WIPO inviabiliza consultas diretas no Google Patents.

Com os identificadores tratados, desenvolveu-se um *web scraper* em Python para extrair das páginas do Google Patents as variáveis necessárias ao treinamento dos modelos de aprendizado de máquina. As variáveis coletadas e calculadas foram:

1. Número de reivindicações independentes;
2. Número de inventores;
3. Número de citações anteriores;
4. Número de imagens da patente;
5. Número de membros da família de patentes;
6. Número de referências não-patente;
7. Número de classificações IPC e CPC;
8. Diversidade tecnológica;
9. Número de substativos no título;
10. Número de aplicações;
11. Soma da quantidade das 10 palavras mais frequentes no titulo na descrição da patetente;
12. Número de citações não patente;
13. Número de diversidade dos IPCs e CPCs.
14. Citações a termo (*forward citations*);

A última variável possui importância especial, pois, de modo análogo aos algoritmos pioneiros de ranqueamento de páginas, a relevância de uma patente pode ser inferida pela quantidade de vezes que é citada.

Para o treinamento dos modelos, utilizou-se um conjunto de patentes publicadas entre 1995 e 2015. Em cada ano, as 10% de patentes com maior número de citações a termo foram rotuladas como "promissoras" (variável-alvo), enquanto as demais foram rotuladas como "não promissoras". Após essa rotulação, a variável de citações a termo foi removida do conjunto de treinamento para evitar vazamento de informação.

A partir do conjunto rotulado, reservou-se 10% para compor o conjunto de teste. Dos 90% restantes, empregou-se uma abordagem semi-supervisionada: 30% permaneceram como dados rotulados e 70% foram tratados como não rotulados. Todo esse processamento foi realizado em Python utilizando a biblioteca Pandas.

Os dados rotulados e não rotulados foram usados para treinar os algoritmos Random Forest e Support Vector Machine (SVM). O Random Forest contribui para mitigar vieses decorrentes do comportamento de uma única árvore, enquanto o SVM fornece uma fronteira de decisão robusta para classificação binária. O desempenho dos modelos foi avaliado no conjunto de teste mediante as métricas AUROC, AUC e F1-Score. O F1-Score permite reduzir falsos positivos e falsos negativos, enquanto a AUROC avalia a capacidade discriminante dos modelos.

O processo semi-supervisionado foi dividido em duas etapas. Na primeira, ambos os modelos foram treinados exclusivamente com o conjunto rotulado, aplicando-se validação cruzada

via GridSearchCV da biblioteca `scikit-learn`. Para o SVM, testaram-se combinações dos parâmetros `C`, `gamma`, `kernel` (mantido como `rbf`) e `class_weight`. Os valores testados foram:

- `C`: 0.1, 1, 10, 100, 1000;
- `gamma`: 1, 0.1, 0.01, 0.001, 0.0001.

Para o Random Forest, os parâmetros avaliados foram:

- `n_estimators`: 100, 200, 300;
- `max_depth`: 10, 20, None;
- `min_samples_leaf`: 1, 2, 4;
- `bootstrap`: True;
- `class_weight`: balanced.

Em ambos os casos, utilizou-se `refit=True` para readequar o melhor modelo ao conjunto completo, `scoring='f1'` para otimizar o F1-Score e `cv=5` para empregar cinco *k-folds*. O GridSearchCV retornou automaticamente os modelos com os melhores hiperparâmetros.

Por fim, na segunda etapa, os modelos treinados foram utilizados para classificar o conjunto não rotulado, e em seguida foram retreinados incorporando essas novas classificações.

5 Resultados

Ao acessar a plataforma do Instituto Nacional da Propriedade Industrial obtivemos um total de 908 patentes em 18 de agosto de 2025 e na plataforma da WIPO obtivemos outro conjunto de 6242 patentes em 10 de outubro de 2025. O conjunto de patentes obtido do INPI possui informações como número do pedido, data de depósito, título da patente e o código da Classificação Internacional de Patentes (IPC). Já o conjunto de patentes obtido do WIPO possui informações como número de pedido, número da submissão, data da submissão, país, título e IPC.

No entanto, só os dados dos dois conjuntos de patentes não são suficientes para o treinamento dos modelos de inteligência artificial. Por essa razão, desenvolvemos um web scraper para complementar os dados. Os dados buscados para cada patente foram: url, título, data de publicação, citações da patente (patent citations), citador por (cited by), resumo (abstract) , descrição (description), quantidade de imagens, documentos similares (similar documents), aplicações que reivindicam prioridade (application claiming priority), eventos Legais(legal events), conceitos (concepts), inventores (authors), outras linguagens (other languages), worldwide applications?, info e links externos (external links). No entanto dado a idade de algumas patentes e a disponibilidade delas na internet ou mesmo no google patets o conjunto final de patentes foi reduzido a 5376 patentes.

Realizando a separação do conjunto de dados total pelas patentes de 1995 a 2015 obtivemos o conjunto de treinamento com 3451 patentes. Deste conjunto obtivemos os valores presentes na tabela 1.

Ano	Promissoras	Não promissoras	Total	Proporção
1995	3	22	25	12.00%
1996	6	50	56	10.71%
1997	10	87	97	10.31%
1998	17	144	161	10.56%
1999	18	156	174	10.34%
2000	17	153	170	10.00%
2001	16	140	156	10.26%
2002	13	110	123	10.57%
2003	14	122	136	10.29%
2004	3	25	28	10.71%
2005	5	42	47	10.64%
2006	7	61	68	10.29%
2007	12	102	114	10.53%
2008	15	129	144	10.42%
2009	22	194	216	10.19%
2010	23	207	230	10.00%
2011	28	249	277	10.11%
2012	37	319	356	10.39%
2013	37	330	367	10.08%
2014	30	262	292	10.27%
2015	22	192	214	10.28%

Tabela 1: Distribuição anual de patentes e proporção de patentes promissoras (1995–2015).

Desse conjunto obtemos 346 patentes para o conjunto de teste, 931 patentes para o conjunto do treino supervisionado e 2174 patentes para o conjunto não rotulado.

Com isso, nós partimos para o treino supervisionado com validação cruzada com o conjunto de 931 patentes e após o fim do treinamento obtivemos os modelos com as seguintes métricas na tabela 2.

aaa	F1-Score	AUROC
SVM	44.83%	86.43%
Random Forest	44.07%	87.05%

Tabela 2: Valores de F1-Score e AUROC dos modelos treinados supervisionadamente.

Usamos cada um dos modelo para classificar o conjunto de 2174 patentes não rotuladas e após a classificação retreinamos o SVM e o Random Forest com o conjunto total rotulado obtendo os modelos finais. Desses modelos obtivemos as seguintes métricas na tabela 3.

aaa	F1-Score	AUROC
SVM	35.29%	85.38%
Random Forest	49.23%	85.61%

Tabela 3: Valores de F1-Score e AUROC dos modelos finais.

A partir dos dados podemos perceber que as métricas caíram em relação aos modelos anteriores. Isso se dá pela probabilidade do modelo anterior não ter aprendido de fato por razões do conjunto de dados rotulado ser pequeno. No entanto, a métrica AUROC continua acima de 50% demonstrando que o modelo aprendeu a diferenciar as patentes promissoras das não promissoras. Mas o F1-Score Decaiu no modelo SVM e continua abaixo dos 50% mostrando que o modelo está classificando patentes promissoras como não promissoras (falsos negativos) e patentes não promissoras como promissoras (falsos positivos). As matrizes de confusão abaixo nas tabelas 4 e 5 demonstram isso.

	Não promissora	Promissora
Não promissora	206	104%
Promissora	5	31%

Tabela 4: Matriz de confusão do modelo SVM.

	Não promissora	Promissora
Não promissora	297	13%
Promissora	20	16%

Tabela 5: Matriz de confusão do modelo Random Forest.

Nós identificamos que o problema do F1-Score pode está no limiar de decisão (threshold) que por padrão é 0.5 nos modelos isso significa que como as probabilidade de uma patente ser promissora é menor que 0.5 dada as proporções nos conjuntos o modelo fica "inseguro" de classificar uma patente com a probabilidade menor que 0.5 de ser promissora causando muitos falsos negativos.

Para resolver esse problema buscamos encontrar um melhor limiar de decisão criando uma função dedicada para encontrá-lo. Um limiar menor 0.5 aumentou (35.29% para 45.28%) o F1-Score do SVM e o limiar do random forest se manteve o mesmo não alterando o F1-Score. Nas tabelas abaixo podemos visualizar o desempenho dos dois modelos.

	Precision	Recall	F1-Score	Support
Não promissora	0.96	0.85	0.90	310
Promissora	0.34	0.67	0.45	36
accuracy		0.83		346
macro avg	0.65	0.76	0.68	346
weighted avg	0.89	0.83	0.85	346

Tabela 6: Relatório de classificação do SVM.

	Precision	Recall	F1-Score	Support
Não promissora	0.94	0.96	0.95	310
Promissora	0.55	0.44	0.49	36
accuracy		0.90		346
macro avg	0.74	0.70	0.72	346
weighted avg	0.90	0.90	0.90	346

Tabela 7: Relatório de classificação do Random Forest.

6 Discussão e Conclusão

A obtenção dos dados representa a parte mais crítica do estudo, pois dados de má qualidade podem comprometer a acurácia do aprendizado dos algoritmos, impedindo que eles encontrem e classifiquem com precisão a "promissoridade" de uma patente. No entanto, a extração dos dados englobou o cálculo de variáveis que exigiram maior complexidade computacional e o acesso a outras fontes de dados que armazenam informações de patentes para complementar os dados extraídos do Google Patents. Outro fator relevante para a busca de dados em outras plataformas foi o fato de algumas patentes mais antigas não estarem disponíveis facilmente na internet. Todos esses fatores fizeram com que a etapa de coleta de dados precisasse de maior alocação de tempo.

Referências

Barros, T. D. (2021). *Etanol* [Acesso em: 20 ago. 2025.]. https://www.embrapa.br/agencia-de-informacao-tecnologica/tematicas/agroenergia/p-d-e-i/etanol

Chung, J., Kim, H., & Lee, S. (2020). Early detection of valuable patents using a deep learning model: Case of semiconductor industry. *Technological Forecasting and Social Change*. https://www.sciencedirect.com/science/article/pii/S0040162520309720

CONAB - Companhia Nacional de Abastecimento. (2021). *Série Histórica das Safras* (Acesso em: 20 ago. 2025.). https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras

Ferreira, A. L. (2009). *Estudo mostra que etanol de cana emite menos gás carbônico para a atmosfera do que a gasolina* [Acesso em: 20 ago. 2025.]. https://www.embrapa.br/busca-de-noticias/-/noticia/18044516/estudo-mostra-que-etanol-de-cana-emite-menos-gas-carbonico-para-a-atmosfera-do-que-a-gasolina

IBM - INTERNATIONAL BUSINESS MACHINES. (2023a). *O que é random forest?* (Acesso em: 20 ago. 2025.). https://www.ibm.com/br-pt/think/topics/random-forest

IBM - INTERNATIONAL BUSINESS MACHINES. (2023b). *O que é uma árvore de decisão?* (Acesso em: 20 ago. 2025.). https://www.ibm.com/br-pt/think/topics/decision-trees

IBM - INTERNATIONAL BUSINESS MACHINES. (2023c). *O que são SVMs?* (Acesso em: 20 ago. 2025.). https://www.ibm.com/br-pt/think/topics/support-vector-machine

INPI - Instituto Nacional da Propriedade Industrial. (2023, dezembro). Boletim mensal de propriedade industrial: estatísticas preliminares [Boletim publicado pela Presidência, Diretoria Executiva, Assessoria de Assuntos Econômicos (AECON).].

Lee, C.-W., Tao, F., Ma, Y.-Y., & Lin, H.-L. (2022). Development of Patent Technology Prediction Model Based on Machine Learning. *Axioms*, *11*(6), 253. https://doi.org/10.3390/axioms11060253

ONU - Organização das Nações Unidas. (2023). *7 - Energia limpa e acessível* (Acesso em: 20 ago. 2025.). https://brasil.un.org/pt-br/sdgs/7

Perrone, C. C., Appel, L. G., Lellis, V. L. M., et al. (2011). Ethanol: An Evaluation of its Scientific and Technological Development and Network of Players During the Period of 1995 to 2009. *Waste Biomass Valor*, *2*, 17–32. https://doi.org/10.1007/s12649-010-9049-z

SEADE - Sistema Estadual de Análise de Dados. (2021). *São Paulo lidera produção de etanol no país* (Acesso em: 20 ago. 2025.). https://informa.seade.gov.br/sao-paulo-lidera-producao-de-etanol-no-pais/

SEBRAE - Serviço Brasileiro de Apoio às Micro e Pequenas Empresas. (2017). *O que é patente?* (Acesso em: 20 ago. 2025.). https://sebrae.com.br/sites/PortalSebrae/ufs/mt/artigos/o-que-e-patente,af88f8ba5a17a510VgnVCM1000004c00210aRCRD

Vian, C. E. F. (2022). *Etanol* [Acesso em: 20 ago. 2025.]. https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/cana/pos-producao/alcool/tecnologias-emergentes/etanol

Zhu, X. (2005). *Semi-Supervised Learning Literature Survey* (rel. técn.). University of Wisconsin-Madison Department of Computer Sciences. http://digital.library.wisc.edu/1793/60444