LegalBERT + SVM + Annoy Architecture Research For Fake News Detection and Explanation Chatbot Project

---

Introduction

Fake news detection has become increasingly crucial in the digital media and digital journalism age. LegalBERT, Support Vector Machines (SVM), and the Annoy (Approximate Nearest Neighbours Oh Yeah) library are combined to form an extremely effective, domain-adaptive pipeline to detect fake news and locate associated evidence-based context. This research outlines the structure, workflow, and advantages of using LegalBERT + SVM + Annoy in the AI-driven "Fake News Detection and Explanation Chatbot" project. This architecture offers a significant advantage when applied to legal domains, particularly news articles focused on politics, court proceedings, and related topics.

LegalBERT:

LegalBERT is a legal domain-specific variant of BERT trained on enormous sets of legal documents, such as court rulings, legislation, and regulations. In the case of spotting fake news, particularly when misinformation is about policy matters, spurious legislation, or legal terminology, LegalBERT works much better than regular BERT models.

Relevance to the Project:

- **Legal and political news**: Numerous fake news pieces use legal terminology to appear trustworthy. LegalBERT understands these semantics more effectively than traditional models.
- **Semantic depth**: Its pretraining on legal texts enhances its ability to manage fact-checking in news focused on regulations or policies

Embedding Generation using LegalBERT:

LegalBERT maps each news article into a high-dimensional, compressed embedding vector that maintains its semantic structure. Both classification (via SVM) and similarity search (via Annoy) are done using these embeddings.

- **Example:** Input news: "A new bill has been passed to legalize free healthcare across all states."
- **LegalBERT output:** Contextual embedding of 768 dimensions per token, pooled for sentence representation.

Support Vector Machine (SVM):

Support Vector Machines are strong supervised learning techniques, especially ideal for binary classification issues such as distinguishing between Fake and Real news

Relevance to the Project:

- **Performance with small datasets**: SVMs perform well even when labelled data is limited.
- **Generalisation**: Better margin separation leads to better generalisation on unseen articles.
- **Explainability**: Feature importance can be derived from the weights assigned to embedding dimensions.

Annoy for Similarity Search:

Annoy is a C++ library that Spotify created with Python bindings for the quick retrieval of approximate nearest neighbours in high-dimensional areas

Relevance to the Project:

- **Contextual evidence retrieval**: Aids the chatbot in retrieving comparable genuine news articles from a reliable database to support the categorisation
- **Low latency**: Suitable for real-time chatbot applications.
- **Lightweight**: Optimized for memory-efficient indexing, allowing implementation on limited infrastructure

Workflow with Annoy:

1. Encode verified real news articles with LegalBERT.
2. Index embeddings using Annoy.
3. For a new search, retrieve top-k similar articles.
4. Use these articles as supporting evidence in the explanation module (Phase 2 of the project).

Specifications of Proposed Pipeline:

Feature	LegalBERT + SVM + Annoy
Model Domain	Legal/Policy-centric
Classification Method	Classic ML (SVM)
Embedding Method	LegalBERT pooled output
Similarity Engine	Annoy (Approximate, fast)
Explainability	Easy via SHAP or weights
Real-time Feasibility	High

Example Use Case:

Input:

"A federal judge ruled that public schools must replace textbooks with TikTok videos by 2026."

LegalBERT + SVM Output:

- Prediction: *Fake*
- Explanation (via Annoy):
    - Similar real news found: "Federal judge allows TikTok usage in classrooms as supplementary tool, not as textbook replacement."
    - Reasoning: Unrealistic mandate, absence from official judicial sources, and no supporting precedent.

Deployment Stack

- **LegalBERT**: via Hugging Face Transformers
- **SVM**: Scikit-learn (Python)
- **Annoy**: Spotify Annoy Python Package
- **Chatbot Interface**: Streamlit / Flask (Phase 2)
- **Explanation Engine**: LIME / SHAP

Summary

The LegalBERT + SVM + Annoy pipeline is a domain-adaptive, scalable, and lightweight transformer-free method that is a good alternative to fully transformer-based approaches. It works extremely well with fake news classification involving legal or policy-based data and ensures user trust through real-time, evidence-based responses. It is thus extremely well-suited for the purpose of the Fake News Detection and Explanation Chatbot project.

References:

1. Chalkidis, I., Androutsopoulos, I., & Aletras, N. (2020). *LEGAL-BERT: The Muppets straight out of Law School* (arXiv:2010.02559). arXiv. https://arxiv.org/abs/2010.02559
2. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html
3. Spotify. (2020). *Annoy: Approximate Nearest Neighbors Oh Yeah* [Computer software]. GitHub. https://github.com/spotify/annoy
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need* (arXiv:1706.03762). arXiv. https://arxiv.org/abs/1706.03762