Review Speak: Predicting Language with Movie Reviews

Predict the next likely word using trigram language modeling on movie reviews corpus

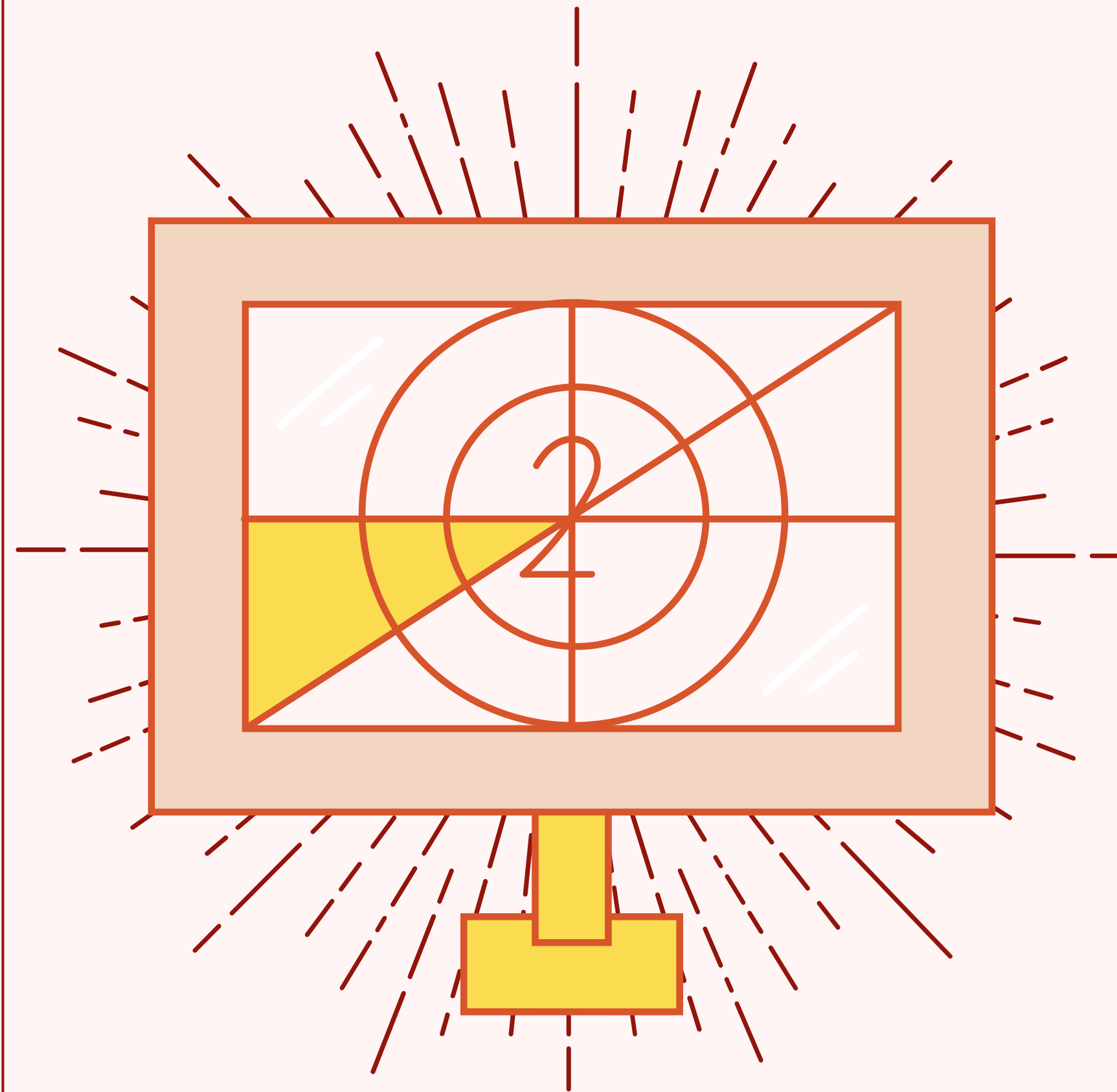EMOTION | AUDIENCE | NARRATIVE | REFLECTION

Presented By:
Olivia Nereparambil

Overview:

- **Build a trigram-based language model using the NLTK movie reviews corpus.**
- **Predict the next most likely word given two input words.**
- **Deploy the prediction interface using Gradio.**

Key Concepts:

- **NLP preprocessing**
- **Trigram modeling**
- **Word prediction**
- **GUI with Gradio**

Libraries & Downloading Resources

Purpose of Libraries:

- **nltk:** NLP processing and corpus access
- **trigrams:** For 3-word pattern generation
- **Counter:** Frequency count of next-word options
- **Gradio**: GUI creation for predictions
- **punkt:** Tokenizer models
- **movie_reviews:** Built-in NLTK dataset with labeled movie review text

Why Needed?
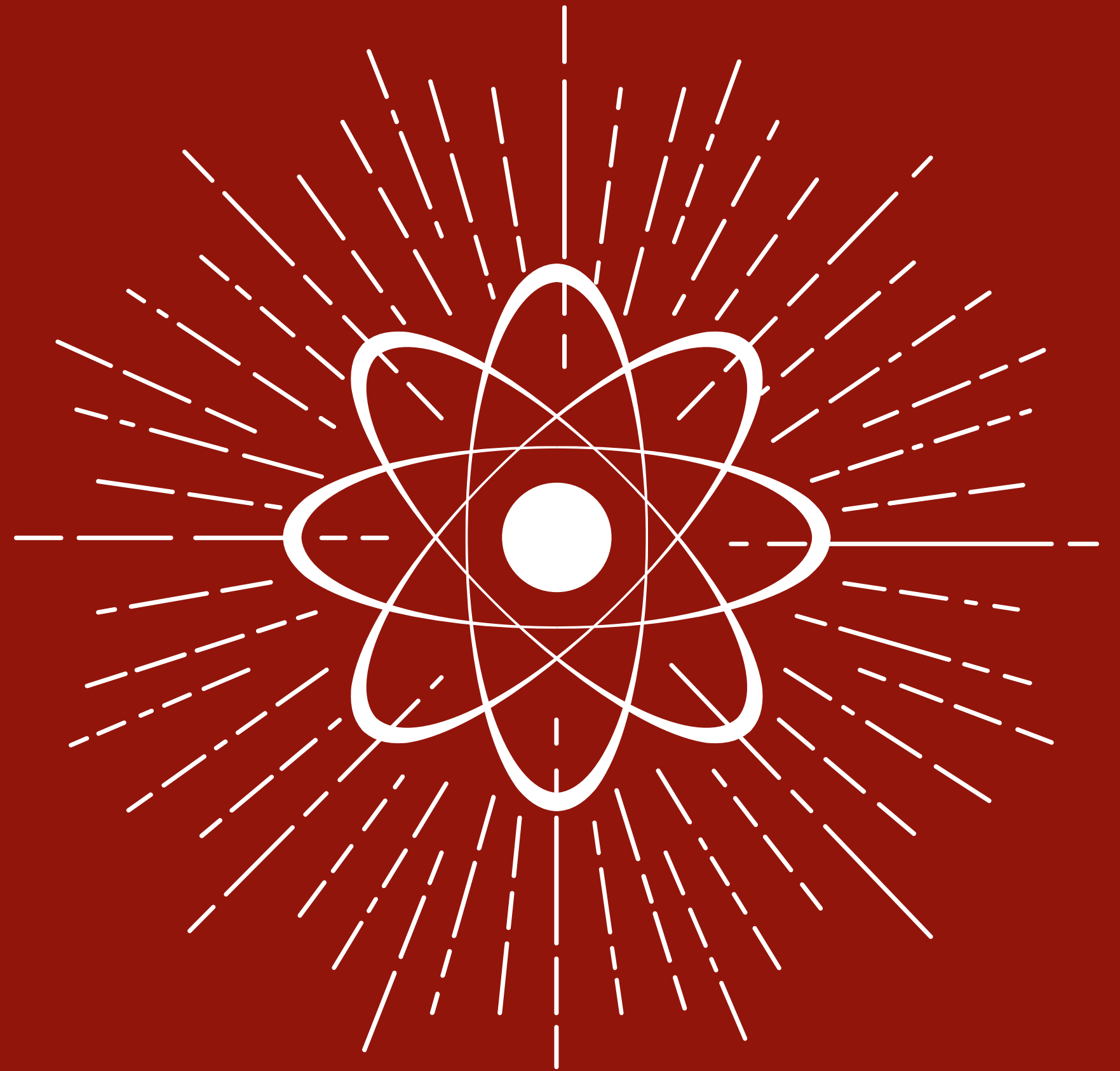**To tokenize and use real-world text for modeling.**

Preprocessing Function

```python
# 🔍 Preprocessing function
def preprocess(text):
    return [word.lower() for word in word_tokenize(text) if word.isalpha()]
```

- **TOKENIZES THE TEXT INTO WORDS.**
- **CONVERTS TO LOWERCASE.**
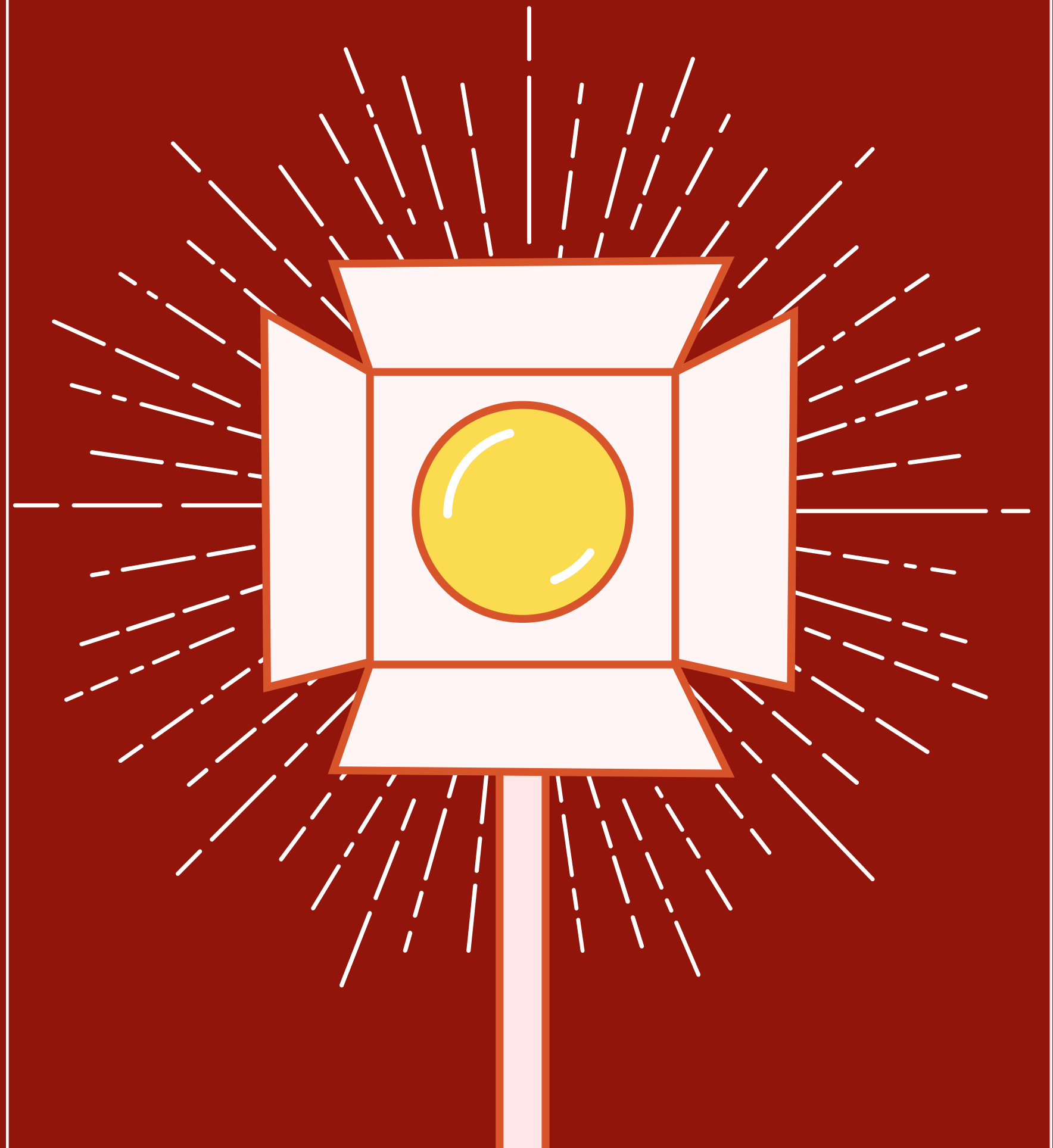- **FILTERS OUT NON-ALPHABETICAL TOKENS (PUNCTUATION, NUMBERS).**

Loading the Corpus & Building the Trigram Model

- COMPILE A FLAT LIST OF ALL CLEANED WORDS FROM ALL MOVIE REVIEWS.
- BUILDS (WORD1, WORD2) → [LIST OF POSSIBLE WORD3S WITH FREQUENCY].
- EXAMPLE: ("THE", "MOVIE") → {"WAS": 10, "IS": 3, ...}

Prediction Function:

- IF THE INPUT BIGRAM EXISTS IN MODEL, RETURN THE MOST COMMON NEXT WORD.
- OTHERWISE, HANDLE UNKNOWN INPUT GRACEFULLY.

GRADIO INTERFACE

A GRADIO INTERFACE IS A WEB-BASED GRAPHICAL USER INTERFACE (GUI) CREATED USING THE GRADIO PYTHON LIBRARY. IT ENABLES DEVELOPERS TO QUICKLY AND EASILY BUILD INTERACTIVE DEMOS AND APPLICATIONS FOR THEIR MACHINE LEARNING MODELS, APIS, OR ANY PYTHON FUNCTION.

ACCEPTS TWO WORDS AS INPUT.
VALIDATES INPUT.
CALLS THE PREDICTION FUNCTION AND RETURNS OUTPUT.

GUI FEATURES:
CLEAN INTERFACE
REAL-TIME PREDICTIONS
INTERACTIVE EXPERIENCE FOR END-USERS

🔮 ReviewSpeak

Enter two English words and get the most likely third word based on movie reviews corpus.

text_input

Enter two words

output

| Clear | Submit | Flag |

Example

Input Words	Predicted Word
the movie	was
this is	not
one of	the
the acting	was

NOTE: RESULTS DEPEND ON PATTERNS SEEN IN THE MOVIE REVIEWS DATASET.

Thankyou