Week 1: Foundations & Preprocessing

Goal: Understand text basics and preprocessing

- **Day 1** → Intro to NLP, tasks, real-world applications
- **Day 2** → Text preprocessing (lowercasing, punctuation, regex cleaning)
- **Day 3** → Tokenization (nltk, spacy)
- **Day 4** → Stopwords removal, stemming, lemmatization
- **Day 5** → POS tagging & Named Entity Recognition (NER) basics (spacy)
- **Day 6** → Hands-on: Build a text cleaning pipeline
- **Day 7** → Mini project: Clean and analyze movie reviews dataset

---

Week 2: Representing Text

Goal: Learn how to convert text into numbers

- **Day 8** → Bag of Words model (CountVectorizer in sklearn)
- **Day 9** → TF-IDF (TfidfVectorizer)
- **Day 10** → Similarity measures (cosine similarity, Jaccard)
- **Day 11** → Word embeddings intro (Word2Vec, GloVe, FastText)
- **Day 12** → Using gensim for Word2Vec
- **Day 13** → Hands-on: Document similarity with embeddings
- **Day 14** → Mini project: Spam email classifier (TF-IDF + ML)

Week 3: Deep Learning for NLP

Goal: Apply neural networks to text

- **Day 15** → Intro to RNN, LSTM, GRU (concept)
- **Day 16** → Text classification with LSTM (Keras/PyTorch)
- **Day 17** → Sequence-to-sequence models (translation basics)
- **Day 18** → Word embeddings with neural nets
- **Day 19** → Sentiment analysis with LSTM/GRU
- **Day 20** → Hands-on: Train LSTM on IMDB reviews dataset

- **Day 21** → Mini project: Build a simple chatbot with seq2seq

---

Week 4: Transformers & Advanced NLP

Goal: Work with modern NLP (BERT, GPT, etc.)

- **Day 22** → Intro to transformers & self-attention
- **Day 23** → Using HuggingFace transformers library
- **Day 24** → Text classification with BERT
- **Day 25** → Named Entity Recognition with BERT
- **Day 26** → Text summarization (extractive & abstractive)
- **Day 27** → Question Answering with pre-trained BERT/GPT
- **Day 28** → Final Project: Choose 1 (Text Summarizer, QA Bot, or Sentiment Analyzer with BERT)

✅ **Outcome**

By the end of this 4-week plan, you'll be able to:

- Preprocess and clean text
- Represent text with BoW, TF-IDF, and embeddings
- Build ML + DL models for text classification
- Use state-of-the-art transformers (BERT, GPT) for real tasks