

Month 4: Natural Language Processing (NLP)

Overview & Schedule (Weeks 13-16)

Week 13: Text Basics

Day	Topic	Theory (Brief)	Practice (Brief)	Time
Day 1	Text Preprocessing	Read: NLP with NLTK (Part 1). Concepts: Tokenization, Stopwords, Stemming vs ...	Python (NLTK/Spacy): Take a raw paragraph. Lowercase, remove stopwords, and t...	2h
Day 2	Vectorization (TF-IDF)	Watch: StatQuest's TF-IDF (10 min). Concept: Frequency * Inverse Doc Frequenc...	Python: Use `TfidfVectorizer` on a list of sentences. Inspect the resulting s...	2h
Day 3	Regex & Parsing	Interactive: RegexOne Tutorial (Lessons 1-10). Pattern matching is crucial fo...	Python: Write a function that extracts all email addresses and dates (dd/mm/y...)	2h

Week 14: Word Embeddings

Day	Topic	Theory (Brief)	Practice (Brief)	Time
Day 1	Word2Vec & GloVe	Read: Jay Alammar's famous Illustrated Word2Vec (20 min read). Concept: 'King...	Python: Load pre-trained GloVe vectors (using `gensim` or `spacy`). Perform v...	2h
Day 2	RNNs for Text	Watch: StatQuest's RNNs Explained (20 min). Concept: Sequential memory, unrol...	Python: Build a simple `nn.RNN` layer in PyTorch. Pass a fake sentence tensor...	2h
Day 3	LSTMs/GRUs	Watch: StatQuest's LSTMs Explained (26 min). Concept: Forget gate, input gate...	Python: Replace your RNN layer with `nn.LSTM`. Output the final hidden state.	2h

Week 15: Modern NLP (Transformers)

Day	Topic	Theory (Brief)	Practice (Brief)	Time
Day 1	Attention Mechanism	Watch: StatQuest's Attention (20 min) OR Read Visualizing Neural Machine Tran...	Conceptual: Draw the Attention matrix. What does it tell us about word relati...	2h
Day 2	BERT & GPT Concepts	Read: Jay Alammar's Illustrated BERT (15 min read). Concept: Masked Language ...	Examine: Look at the architecture diagrams. Note the difference between Encod...	2h
Day 3	Hugging Face Library	Watch: Hugging Face's Transformers Library Tour (10 min). Concepts: `from_pre...	Python: Use `pipeline('sentiment-analysis')` to classify 5 sentences with zer...	2h

Week 16: NLP Project

Day	Topic	Theory (Brief)	Practice (Brief)	Time
Day 1	Project: Text Classification	-	Fine-tune a model for IMDb reviews.	2h
Day 2	Pipeline Integration	-	Build a clean inference function.	2h
Day 3	Review	-	Compare TF-IDF vs BERT performance.	2h