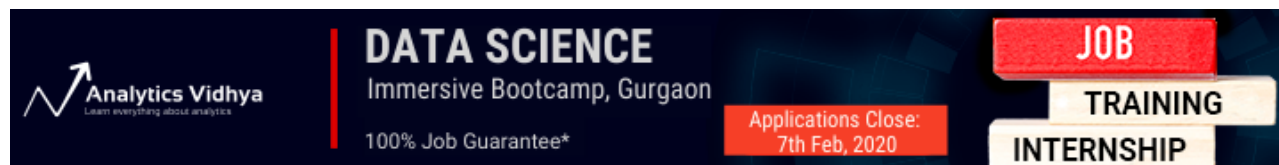


A Comprehensive Learning Path to Understand and Master NLP in 2020

ANALYTICS VIDHYA, JANUARY 8, 2020 [LOGIN TO BOOKMARK THIS ARTICLE](#)



The banner features the Analytics Vidhya logo on the left. The main text reads 'DATA SCIENCE Immersive Bootcamp, Gurgaon' with a '100% Job Guarantee*' below it. A red box on the right says 'Applications Close: 7th Feb, 2020'. On the far right, a vertical stack of three boxes contains the words 'JOB', 'TRAINING', and 'INTERNSHIP'.

Introduction

Google “NLP jobs” and a remarkable number of relevant searches show up. There are businesses spinning up around the world that cater exclusively to Natural Language Processing (NLP) roles! The industry demand for NLP experts has never been higher – and this is expected to increase exponentially in the next few years.

But the supply side of things is falling short. Freshers and even experienced folks who want to land an NLP based role are struggling to break into the industry. We can pinpoint one of the biggest pain areas – a lack of structured learning.

There are far too many resources these days that cover NLP concepts but the majority of these do so in a scattershot manner. Freshers tend to pour through articles and books, parse various blogs and videos, and end up struggling to piece together an end-to-end understanding.

This is where our NLP learning path comes in! We are thrilled to present a comprehensive and structured learning path to help you learn and master NLP from scratch in 2020!



This learning path has been curated by experts at Analytics Vidhya who have gone through hundreds of resources to curate this for our community. Follow this path in 2020 and you'll be on the verge of landing a role in the NLP domain soon!

Our Framework for the NLP Learning Path

Structure – that's at the heart of everything we do. Our learning paths are popular for their structure as well as their comprehensive nature. Here's how we've broken down each month of the NLP learning path to help you plan your learning journey:

- **Objective:** What will you learn in that month? What are the key takeaways? How will your NLP journey progress? We mention this at the start of each month to ensure you know where you stand and where you will be at the end of that particular month
- **Time Suggested:** How much time on average you should spend on that section per week
- **Resources to Learn:** A collection of the top resources for the NLP topics you will learn in that month. This includes articles, tutorials, videos, research papers, and other similar resources

Looking for other learning paths in data science? Your wait is over:

- [The Learning Path to become a Data Scientist and Master Machine Learning in 2020](#)
- [The Learning Path to Master Deep Learning in 2020](#)
- [Computer Vision Learning Path \(Launching January 9th\)](#)

Let's dive into it!

Month 0 – Prerequisites (Optional)

Objective: This is for all of you who are not yet familiar with Python and Data Science. By the end of this month, you should have a fair idea about the building blocks of machine learning and how to program in Python.



Time Suggested: 6 hours/week

Python for Data Science:

- [Course: Python for Data Science](#)
- [Python Cheat Sheet](#)

Learn Statistics:

- [Descriptive Statistics](#) by Khan Academy

Data Preparation:

- Training and Testing:
- [Split Data using sklearn](#)

Linear Regression:

- [A Comprehensive Guide on Linear Regression](#)
- Video on Linear Regression:

Logistic Regression:

- [Logistic Regression using Python](#)
- Video on Logistic Regression:

Decision Tree Algorithm:

- [Tutorial on Tree-Based Algorithms](#)
- Introduction to Decision Trees:

K-fold Cross-Validation:

- [Improve Your Model Performance using Cross-Validation \(in Python and R\)](#)
- K-Fold Cross Validation Video:

Singular Value Decomposition (SVD):

- [SVD from Scratch](#)
- SVD by Gilbert Strang:

Month 1 – Getting Comfortable with Text Data

Objective: And off we go! This month is all about getting you familiar and comfortable with the basic text preprocessing techniques. You should be able to build a text classification model by the end of this section.



Time Suggested: 5 hours/week

Load Text Data from Multiple Sources:

- [Pandas: How to Read and Write Files](#)

Learn to use Regular Expressions:

- [Basics of Regular Expressions](#)
- [Speech and Language Processing](#) by Stanford

Text Preprocessing:

- [spaCy library](#)
- [Tokenization using the spaCy library](#)
- [NLTK Library](#)
- [Stopword Removal and Text Normalization](#)

Exploratory Analysis of Text Data:

- [A Complete Exploratory Data Analysis and Visualization for Text Data](#)

Extract Meta Features from Text:

- [Traditional Methods for Text Data](#)

Project:

- Build a Text Classification model using Meta Features. You can use the dataset from the practice problem [Identify the Sentiments](#)

Month 2 – Computational Linguistics and Word Vectors

Objective: This month you will start to see the magic of NLP. You will learn how English grammar can be utilized to extract key information from text. You will also work with word vectors, an advanced technique to create features from text.



Time Suggested: 5 hours/week

Extract Linguistic Features:

- Part-of-Speech Tagging using spaCy:
- Named Entity Recognition using spaCy:
- Dependency Parsing by Stanford:

Text Representation in Vector Space:

- [Bag of Words, TF-IDF and Word Embeddings](#)
- Word Embeddings:
 - Word Vector Representations (Word2Vec) by Stanford:
 - [Text Classification & Word Representations using FastText](#)
 - [Tool: Gensim – Word2Vec](#)

Topic Modeling:

- [Topic Modeling using Latent Semantic Analysis](#)
- [Beginner's Guide to Topic Modeling in Python](#)
- Topic Models:

Information Extraction:

- [Information Extraction using Python and spaCy](#)

Projects:

- Build Sentiment Detection Model using Word Embeddings. You can use the dataset from the practice problem [Identify the Sentiments](#)
- Categorize News Articles using Topic Modeling

Month 3 – Deep Learning Refresher for NLP

Objective: Deep learning is at the heart of recent developments and breakthroughs in NLP. From Google's BERT to OpenAI's GPT-2, every NLP enthusiast should at least have a basic understanding of how deep learning works to power these state-of-the-art NLP frameworks. So this month, you will focus on the concepts, algorithms, and tools around Deep Learning.



Source: Tryolabs

Time Suggested: 5 hours/week

Neural Networks:

- [Introductory Guide to Deep Learning and Neural Networks](#)

Optimization Algorithms:

- [Optimization Algorithms for Deep Learning](#)

Recurrent Neural Networks (RNNs) and LSTM:

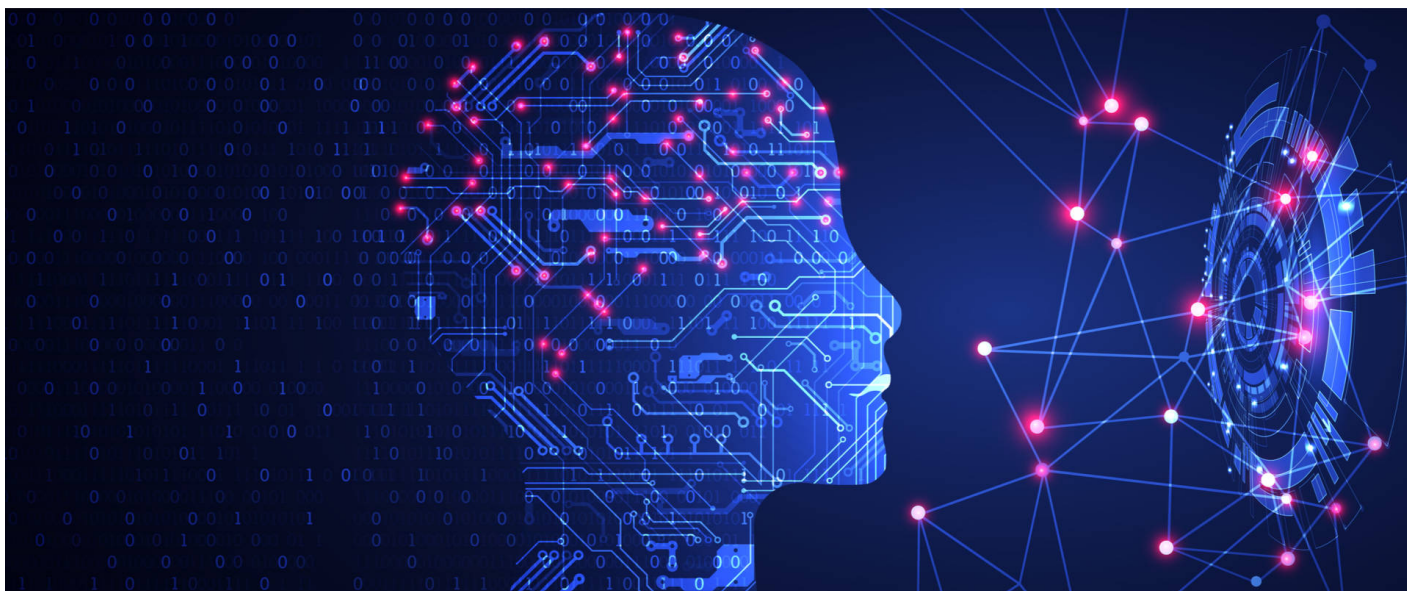
- A friendly introduction to RNNs:
- [Recurrent Neural Networks Tutorial, Part 3 – Backpropagation Through Time and Vanishing Gradients](#)
- [Research Paper: Fundamentals of RNN and LSTM](#)

Introduction to PyTorch:

- [A Beginner-Friendly Guide to PyTorch](#)
- [Course: Introduction to Deep Learning with PyTorch](#)

Month 4 – Deep Learning Models for NLP

Objective: Now that you have a taste of deep learning and how it applies in the NLP context, it's time to take things up a notch. Dive into advanced deep learning concepts like Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM), among others. These will help you gain a mastery of industry-grade NLP use cases.



Time Suggested: 5 hours/week

Recurrent Neural Networks (RNNs) for Text Classification:

- [RNN – PyTorch](#)
- [Sentiment Analysis using LSTM](#)
- [Understanding Bidirectional RNN in PyTorch](#)

CNN Models for NLP:

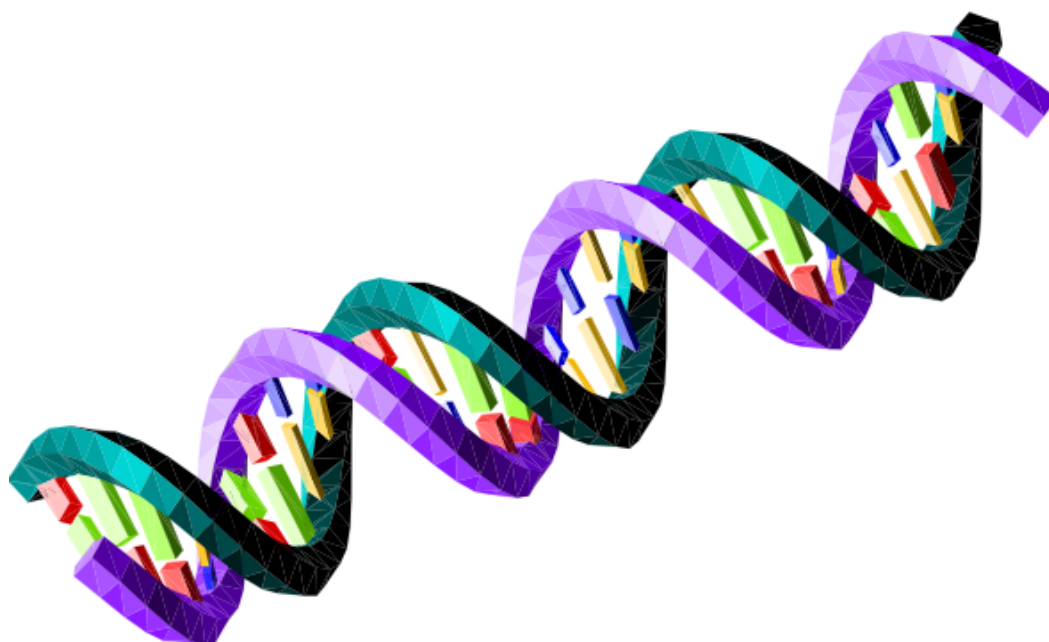
- [Understanding CNN for NLP](#)

Projects:

- Build a model to find named entities in the text using LSTM. You can get the dataset from [here](#)

Month 5 – Sequential Modeling

Objective: In this month, you will learn to use sequential models that deal with sequences as inputs and/or outputs. A very useful concept in NLP as you'll soon discover!



Time Suggested: 5 hours/week

Language Modeling:

- Language Models and RNNs by Stanford:
- [A Comprehensive Guide to Build your own Language Model in Python!](#)
- [Text Generation with PyTorch](#)
- [Research Paper: Regularizing and Optimizing LSTM Language Models](#)
- [Book: Speech and Language Processing – N-gram Language Models](#)

Sequence-to-Sequence Modeling:

- [PyTorch Seq2Seq](#)
- [Research Paper: Sequence to Sequence Learning with Neural Networks](#)
- [Seq2Seq with Attention](#)

Projects:

- Train a language model on [Enron Email dataset](#) to build an auto-completion system
- Build a Neural Machine Translation Model (English to any language of your choice)

Month 6 – Transfer Learning in NLP

Objective: Transfer learning is all the rage in NLP at the moment. This has actually helped democratize the state-of-the-art NLP frameworks you would have come across before. This month introduces BERT, GPT-2, ULMFiT and Transformers.



Time Suggested: 5 hours/week

ULMFiT:

- [Text Classification using ULMFiT in Python](#)
- ULMFiT by FastAI:

Transformers:

- [Research Paper: Attention is all you Need](#)
- [How do Transformers Work in NLP?](#)

Pre-trained Large Language Models (BERT and GPT-2):

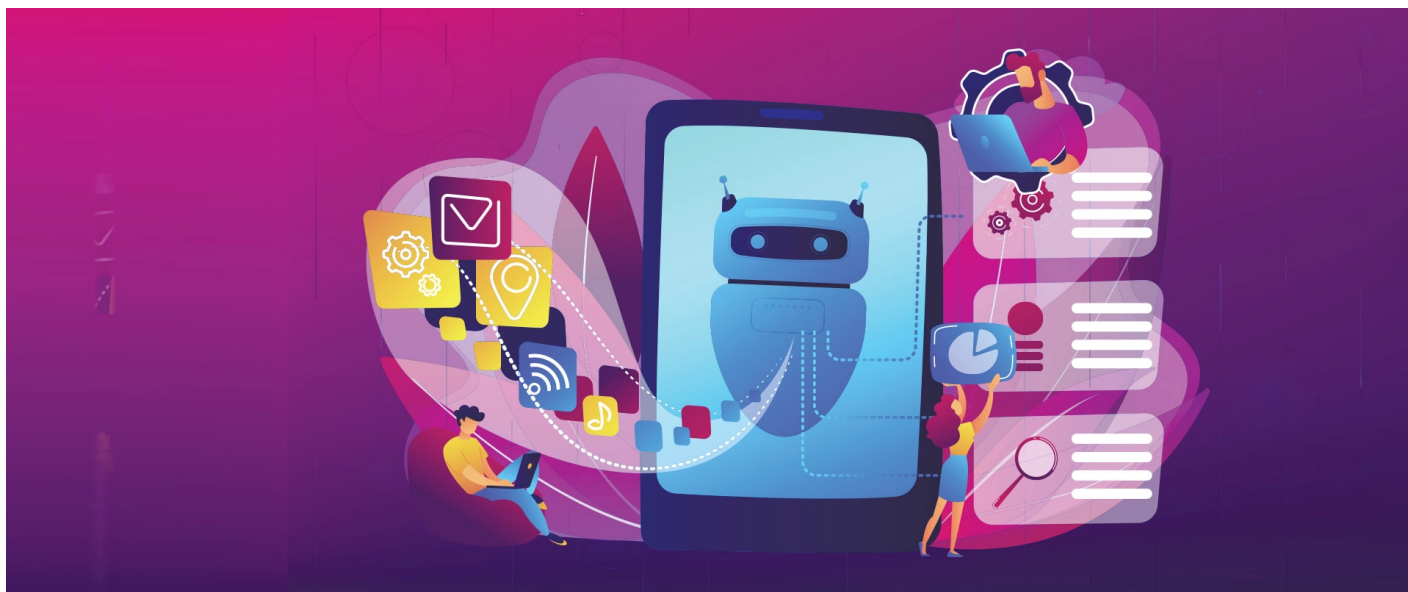
- [Demystifying BERT](#)
- [Bert-As-a-Service](#)
- [Research Paper: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#)
- [Tool: Transformers](#)

Fine-Tuning pre-trained Models:

- [BERT Fine-Tuning Tutorial with PyTorch](#)

Month 7 – Chatbots and Audio Processing

Objective: You will learn how to build a chatbot or conversational agent this month. Once you have mastered NLP, the next frontier you can tackle is Audio Processing.



Time Suggested: 5 hours/week

Chatbots:

- Rasa Masterclass:
- [Learn how to Build and Deploy a Chatbot in Minutes using Rasa](#)
- [How to build a voice assistant with open source Rasa and Mozilla tools](#)

Audio Processing:

- [Speech Data Exploration](#)
- [Audio Classification](#)
- [Pre-trained speech-to-text model – DeepSpeech](#)

Project:

- Build a chatbot with voice interface using Rasa

Infographic – NLP Learning Path for 2020

Our community loves the infographics we design for each learning path. These infographics serve two primary purposes:

- They help us visualize the structure of how we'll learn different topics

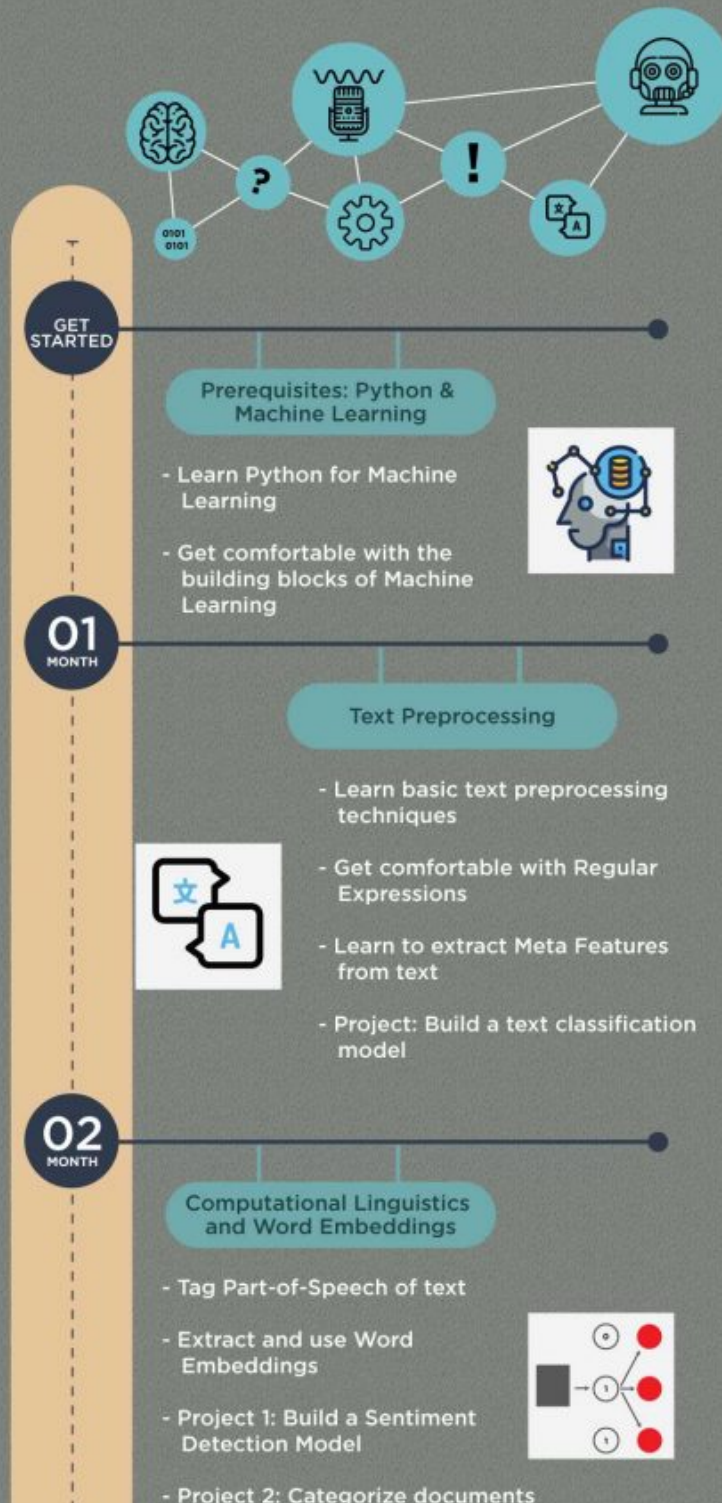
- They can be used as checklists to tick off concepts as you progress in your NLP journey

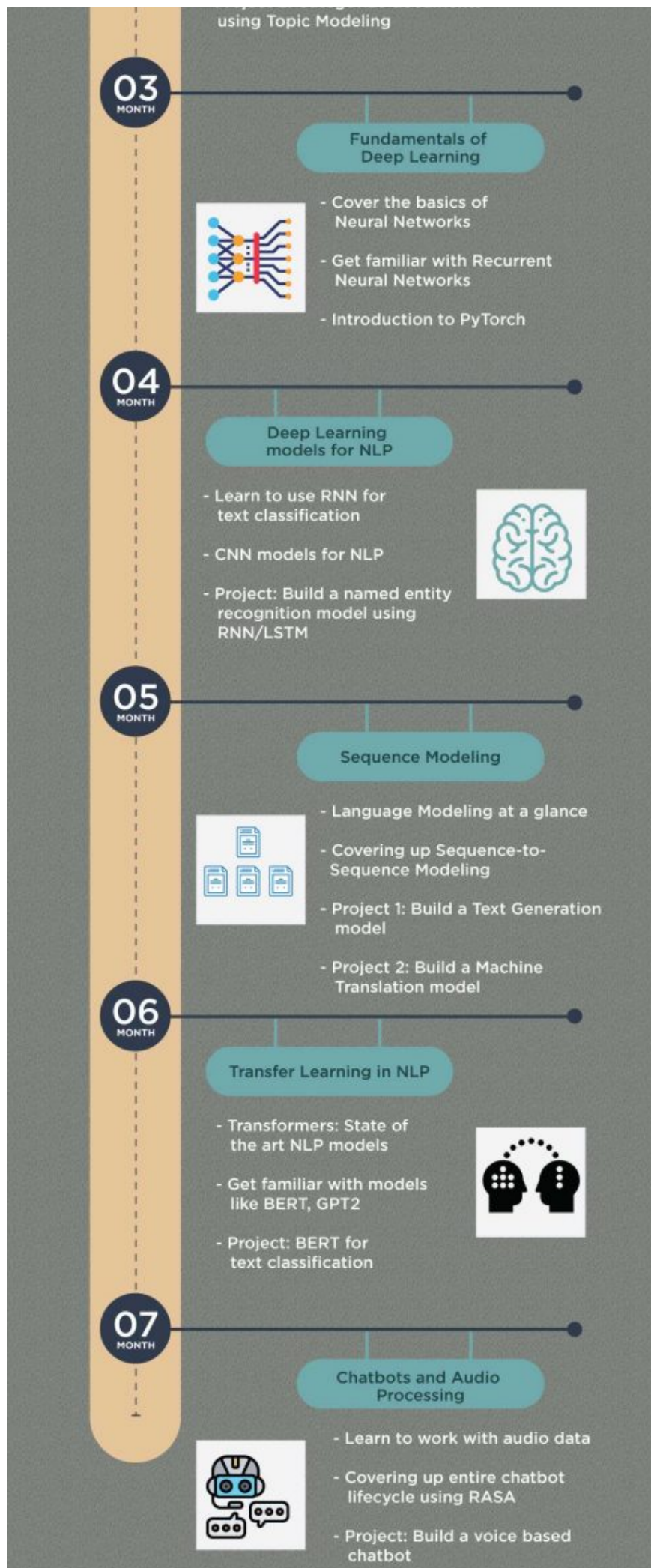
So, we're thrilled to present below the NLP learning path infographic for 2020! *You can download a high-resolution version from [here](#).*

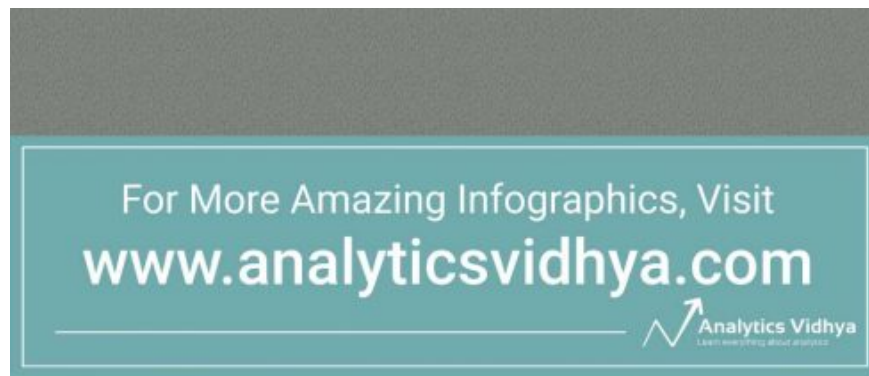
LEARNING PATH 2020

NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is the hottest field to work in right now. Here is a comprehensive step-by-step learning path to master NLP in 2020! We have curated the best NLP learning resources to enable your inner NLP expert this year.







You can also read this article on Analytics Vidhya's Android APP



Share this:

Related Articles



[Your Ultimate Learning Path to Become a Data Scientist and Machine Learning Expert in 2020](#)

January 6, 2020

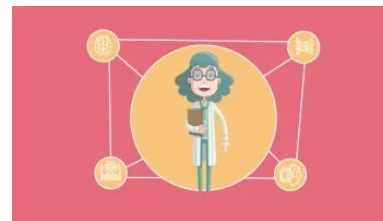
In "Analytics Vidhya"



[2019 In-Review and Trends for 2020 - A Technical Overview of Machine Learning and Deep Learning!](#)

December 23, 2019

In "Beginner"



[A Comprehensive Learning Path for Deep Learning in 2020](#)

January 7, 2020

In "Beginner"

TAGS : [LEARN NLP](#), [LEARNING PATH 2020](#), [NATURAL LANGUAGE PROCESSING](#), [NATURAL LANGUAGE PROCESSING LEARNING PATH](#), [NLP](#), [NLP CONCEPTS](#), [NLP EXPERT](#), [NLP LEARNING PATH](#), [NLP PROJECTS](#)

NEXT ARTICLE

Here's your Learning Path to Master Computer Vision in 2020

PREVIOUS ARTICLE

A Comprehensive Learning Path for Deep Learning in 2020



[Analytics Vidhya](https://www.analyticsvidhya.com)

This is the official account of the Analytics Vidhya team.