


Varun CHITALE

in [linkedin.com/in/varun-chitale](https://www.linkedin.com/in/varun-chitale)  github.com/varunchitale
📞 +91 8698 471979 @ varunchitale@gmail.com
🌐 Website 📍 Pune, India





Data Scientist | NLP Engineer

I am currently working in the legal-tech field, solving all kinds of Natural Language Processing tasks using sophisticated Machine Learning approaches. These tasks include- classification of text, clustering, generating summaries- abstractive and extractive, working with chatbots, topic modelling, named entity recognition and building production ready ML pipelines. I have extensive experience working with Postgres, Python and its frameworks, and AWS. I like publishing programming snippets that I find interesting on my  [Medium blog](#). I enjoy watching Seinfeld, spending hours tuning in to the tunes for Pink Floyd, and hiking.

SKILLS

Programming Languages	Python (Spacy, Tensorflow, ScikitLearn, Flask, Keras, Cython, nltk, matplotlib, seaborn), R, SQL, C, C++, Octave
Databases	MySQL, PostgreSQL
Postgres	Stored procedures, Indexing, Triggers, Notification Systems, Views, PL/pgSQL
Development Tools	Sublime, Eclipse, git, VSTS, Docker
Operating Systems	Linux(Arch Linux, Ubuntu), Mac OS X, Windows
Others	AWS EC2, AWS Lambdas, VNC, AWS Rekognition, RESTful APIs, \LaTeX , SSH

PROFESSIONAL EXPERIENCE

Present June 2018	Data Scientist, NLP,  Riverus Technology Solutions <ul style="list-style-type: none">> Translating stakeholder requirements into achievable tasks(within the NLP and Data Science domain) and setting up timelines> Setting up a real-time orchestrated NLP pipelines using Flask, uWSGI, triggers and servers> Organizing data quality reports> Fine tuning of existing models> Organizing annotated data and storing it in the db> NLP Search, Semantic Search, Autocomplete  Medium Article> A guide on Mergers and Acquisitions within the Indian Tax Domain> QA system for legal question answering <div>Python R Postgres AWS PowerBI NLP ML Pipelines Data Analytics Data Visualization</div>
November 2017 May 2018	Associate, DATA SCIENTIST, Riverus Technology Solutions <ul style="list-style-type: none">> Database Architecture for storing varied NLP data> NER Models for extraction of meta data> Topic Modeling and clustering similar legal judgments> Classification of text in a legal judgment to extract main issues of the case, reasoning and outcome elements  Medium Article> Data Validation Tests> Basic accuracy reports <div>Python Flask Postgres MySQL ScikitLearn Keras Model Optimization In house Annotation</div>
October 2017 August 2017	Intern, DATA SCIENTIST, Riverus Technology Solutions <ul style="list-style-type: none">> Understanding fundamental concepts of Data Science in NLP> Sentence Tokenization, Word Tokenization- Maximum Entropy Models> Extraction of various linguistic elements from legal documents <div>Python JSON ScikitLearn Data Extraction</div>
2016	Web Developer, DATA ANALYST,  Biomark Labs <ul style="list-style-type: none">> Data Analytics and Insights over 2000+ files.> Worked on setting up the database architecture.> Implemented Apriori algorithm to improve product catalog listing.> Built a website powered by a MySQL db. <div>PHP R MySQL CRM</div>

2013-2017 **MIT, College of Engineering, University of Pune**
 Bachelor of Computer Engineering
 Score : **81/100** Dept. Rank : **1/142**
Selected Coursework : Business Analytics and Intelligence, Data Mining Techniques and Applications, Smart System Design and Applications. Engineering Mathematics I-III — Differential and Integral Calculus, Measure of Central Tendency and Variance, Vector Algebra, Fourier and Taylor Series, etc.

PROJECTS AND RESEARCH WORK

A DISCOURSE ON PRECEDENTS WITHIN THE INDIAN INCOME TAX DOMAIN

2017 - 2018

 [Request Report](#)

A legal case has various subversive properties which can be understood if its precedents(or cases it refers to) are studied carefully. This discourse outlines the history of precedents, the current state of affairs, various doctrines, ways of extracting and storing this information from legal documents, and in-depth analysis of the aforesaid claim.

Law NER Spacy Stanford NER CoreNLP Python R

RESEARCH : A NOVEL APPROACH TO AUTOMATIC TEXT SUMMARIZATION.

2016 - 2017

 [Request Report](#)

We propose an unsupervised algorithm for generating extractive summaries domain-free text. This was packaged in a standalone GUI app built entirely in Python using Tkinter. This was my final year project. Appreciated by Persistent Systems, India and lauded as the best project at school.

Python Tkinter Text-Rank Algorithm Unsupervised Learning

PROJECT SEVERUS

2017 - 2018

 [Riverus](#)

A fully functional web app built using React/Redux and Python to automate everything. Provides features for annotations used in all of the in-house ML tasks, logging user behaviour to improve searches, real time execution of ML pipelines, organizational dashboards and reports and so on. A tool that focuses on the other end of the telescope- the annotators and the stakeholders

React/Redux Python Flask AWS JWT Tensorflow Spacy Postgres

LANGUAGES

English	● ● ● ● ●
German	● ● ● ● ○
Hindi	● ● ● ● ●
Marathi	● ● ● ● ●
French	● ● ○ ○ ○

MUSIC

- > **Guitar**
Trinity School of London
 - > Grade 3, Rock and Pop
 - > Grade 2, Plectrum
- > **Tabla** (percussion instrument comprising of two small drums)
Bharat Gayan Samaj
 - > Grade 2, Second class

REFERENCES

Rutuja Udyawar
 Statistician and VP Analytics
Riverus Technology Solutions
 @ rutuja@riverus.in
 ☎ +91 99229 63860

Sheece Baghdadi
 COO
Riverus Technology Solutions
 @ sheece@riverus.in
 ☎ +91 98203 38869