Internship Report

**Intern Details:**

|  |  |
| --- | --- |
| **Student Name** | **RAJANA VEDA PRADEEP nAIDU** |
| **Roll Number** | **CS15B054** |
| **Company of Internship** | **eclerx** |
| **Profile** |  |
| **Internship start date** | **17-5-2018** |
| **Internship duration** | **2 months** |
| **Location** | **Mumbai** |
| **Designation** | **XXXXXXX** |
| **Internship Manager** | Mr. NITIN THAKKAR |
| **Internship PM Coach** | **XXXXXXX** |
| **Project Name** | **XXXXXXX** |
| **Team** | **XXXXXXX** |

**Project Overview:**

Name of the project: **TEXT aLIGNMENT**

Languages used: **r , Python**

Frameworks used: **r Studio, Py charm**

Environments used: **R Studio ,Anakonda**

Tools/Services used: ***Wordnet , datamuse,Tesuras, big .txt taken from Norvig.com***

Hosting Server used:

*Libraries used in R code:*

* rvest ,stringi ,stringr,stringdist, ggplot, csv,xml2,
* system libraries like utils

*Libraries used in Python code:*

* numpy ,scipy ,scikit-learn , tensorflow , keras

Extent of Skills used:

* **XXXXXXX**
* **XXXXXXX**
* **XXXXXXX**
* **XXXXXXX**

**Project Details:**

* Salient Features:
* Text Alignment in R ,Word prediction using Python
* Part1(Code in R) : a)Creating data for corpus file , sorting the words in decending order of frequency . b) correctin of word is based on edit to a word is a deletion (remove one letter) or a replacement (change one letter to another) or an insertion (add a letter) or a transposition (swap two adjacent letters) . c)Based on the sorted words taken from the big corpus file ,we apply this algorithm to correct the given word .
* Part2(Code in R) : a)The above word correction is extended to line correction & then to multiple lines ,Hence able to to text alignment of all the word in the text ,word by word . b)This thing is further extended to user defined words , user defined has higher priority than the sorted corpus file , while aligning word
* Part3(Code in R) : a) web scraping in R , while grouping the given set of words into word class , to form the word classes , scraped the dynamic data in R .
* Part4(Code in Python): a) word2vec implementation using Python , using 2 different models with front end keras & Back end tensorflow(tf)

**Learning of the student:**

***Technical:***

* Got good exposure to R language and R studios
* Learned Algorithims and statics part that r related to text analysis
* In the period of Internship got good understanding of Natural Language Processing(NLP)
* Got to know more about about Keras and Tensorflow and CNKT . Implemented word2vec representation using (continuous bog of words)CBOW & n-skip grams models for predicting contex & labels.

***AESTHETIC****:*

* Good modeling of code
* **XXXXXXX**
* **XXXXXXX**
* **XXXXXXX**

**RATIONALE:**

Software Development Life Cycle (SDLC) learnings:

* **XXXXXXX**
* **XXXXXXX**
* **XXXXXXX**
* **XXXXXXX**

Personal Learnings:

* This Internship helped me to apply my knowlege in computer Science and got explosure to more things that are related to CSE .
* **XXXXXXX**

**Non-revelation of confidential Information**

**XXXXXXX**