|  |  |
| --- | --- |
| Internship Project Title | Automate detection of different emotions from paragraphs and predict overall emotion |
| Project Title | Emotion Detection From text |
| Name of the Company | TCS iON |
| Name of the Industry Mentor | Soumyadip Mal |
| Name of the Institute | IIT Dharwad |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 21st May 2020 | 6th July 2020 | ~130 hrs | Both  Google Colab  as well as  Jupiter Notebook | Numpy ,  Pandas,  NLP(ntlk),  Sklearn |
| Project Synopsis:  1. Aim is Automate detection of different emotions from paragraphs and predict overall emotion  2. firstly this should be classification type of ML model  2. then prepossessing of data & data cleaning is Done  3. Vectorization of data to enhance the speed by taking multiple iterative operations among data points and turning them into matrix operations  4. Then fitting this data with suitable ML model with good accuracy  5. Predication for any new input | | | | |
| Solution Approach:  1. this will be classification based ML model  2. so firstly converting data into pandas-dataframe  3. then cleaning this data by removing/replacing stop words,URLs, Emojis, punctuation marks & replacing short-forms  4. removing irrelevant data  5. as no fixed features are used in case of such classification, vectorization of data with different ways(countVectorizer/TfidfVectorizer) will be helpful  6. as of classification ML : we need to try different classification ML model according to accuracy of model we should accept it | | | | |
| Assumptions:  1. as there are data is unbalanced & many of emotions are quite similar creates lot or error in data-set  2. so merging such closely related emotions overcome such small error  such as : Anger / Boredom / Hate = Disgust  3. Neutral emotion can be titled over any emotion but this creates more chaos for any ML model leading in drop of accuracy so in such case dropping of such class completely is always better for improving performance of model  e.g.  for data-set considered have neutral emotion for following which makes no sense  neutral → im saaaaaaaaaaaaaad i need someone to talk to.  neutral → lonely ..sad | | | | |
| Project Diagrams:  wow let's drive the car → enthusiasm  i want to hit you → angry  i got new pencil → happy  Noway ! India lost the worldcup → sad    yeah i got new macbook → ????  yeah i got new macbook → happy | | | | |
| Algorithms:  sgdclassifier :  Stochastic Gradient Descent. Stochastic Gradient Descent (SGD) is a simple yet very efficient approach to fitting linear classifiers and regressors under convex loss functions such as (linear) Support Vector Machines and Logistic Regression. | | | | |
| Outcome:  SGDClassifier  TfidfVectorizer  accuracy\_score : 0.6172968145281333  f1\_score : 0.5536703949888738  precision\_score : 0.6039685467932462  this ML model gives accuracy upto 61.73% | | | | |
| Exceptions considered:  there are many emotions, which are quite similar in such cases strict classified data-set is not possible which leads reductions of accuracy  so in such case grouping such emotions under particular emotion improves accuracy  e.g.  Anger / Boredom / Hate = Disgust | | | | |
| Enhancement Scope:  improved and properly classified dataset may increase the accuracy  in general just small change in tone, changes the meaning of same sentence. So For analysis emotions of some paragraphs just text is not enough | | | | |
| Link to Code and executable file:  Google colab :  https://colab.research.google.com/drive/1NQveSuvkc2SrJhmzwpWak1isF0XAb-AB?usp=sharing  GitHub :  <https://github.com/Shiru99/Emotion_Detection> | | | | |