Kartik k karande

Phone No - 9527722700 Gmail- karandekartik777@gmail.com |

LinkedIn - https://www.linkedin.com/in/kartik-karande-5333391ab/

Profile

An independent and self-motivated student with proven ability in developing Data science projects, strong apt for learning new skills curious to grow in dynamic and fast paced workplace. I have also participated in various hackathons and secured success. Want a challenging carrier as Data scientist where Python Machine Learning skills can be effectively used and upgraded.

Achievements

Published a Research paper on Cybersecurity using Machine learning
https://www.ijcaonline.org/archives/volume174/number31/31876-2021921249
Winner of National level hackathon (PCMC Smart city hackathon 2020)

Winner of hackathon (symbiosis institute of technology)

Winner of National level Mini ROBOCON 2019

Running a YouTube channel for python and Data science
https://www.youtube.com/watch?v=NFefvBfpn8s

Written a Blog on Transformers
https://github.com/karandekartik777/Blog-on-Transformer

Experience

- **Proven ability of experience in developing Data science and machine learning models as a Jr. Data scientist**
- **Working as a Deep learning intern At ineuron.Ai**

Functional Responsibilities

- involved in Data Preprocessing Techniques for making Data useful for creating Machine learning Models
- Creating Machine learning models and tune then to achieve good performance
- Translate product requirements into analytical requirements
- Lead and Mentor a team of students in my final year .
- Involved in creating various regression classification, clustering algorithms by using various sklearn libraries
- Sharing Blogs on NLP and Data science

Other skills

- **Self-Learner and Proficient at Googling and reading documentations**
- **Know project Development life cycle**

Technical skills: **python , C++ , Data science , Machine learning , Deep Learning, Natural language processing , Exploratory Data analysis, Feature Engineering , Data visualization ,Regression algorithms , clustering algorithm , Bagging ,Boosting , Statistics, dimensionality reduction , Neural Network, CNN, RNN, LSTM, Attention based models and transfer learning**

Libraries : **Scipy,Numpy , pandas ,seaborn ,sklearn , plotly ,matplotlib ,NLTK , Xgboost Spacy ,genism ,cufflinks ,TensorFlow, keras , PyTorch ,pyspark , Time series , LSTM ,Transformers ,BERT ,spark mlib ,xgboost , vgg16, Resnet**

Courses and Certifications:

- **Introduction to Data science (Coursera)**
- **Machine learning A to Z (Udemy)**
- **Deep learning specialization (Coursera)**
- **Deep learning NLP Masters (Ineuron.ai)**
- **Introduction to Data science (Coursera)**

Projects

1.Quora question Pair similarity:

Quora is the application where users can answer question and experts can give answers , Identify which questions asked on Quora are duplicates of questions that have already been asked This could be useful to instantly provide answers to questions that have been already answered. But users ask many questions which are repeated or someone asked it before , or they can be asked in different wording . hence it is time consuming and hectic to the expert who are going to answer . so we are created a ML model which can find the text similarity and let us know that incoming question is repeated or not if it is repeated then the answer of the same question which is asked before will be send to user and load on expert or answer giver will be reduced.

ROLE :

1)gathered Data from Kaggle **2)** perform EDA **3)** text and Data preprocessing (Stopwords removal , creating tokens , word to vector representation) **4)** Feature Engineering (created new features like Common word counts, common word share , question 1 and question 2 length cosine similarity, Minimun edit distance) **5)**created features like wordvector **6)** created a random model and check performance **7)**created machine learning models like Logistic regression ,Random forest and Xgboost **8)**Use log loss as performance metric

2.Linkous (Detection of Malicious links and URL'S using Machine learning)

(Final year project) http://linkous.herokuapp.com/#/

In this project is related to cybersecurity and machine learning and NLP . in this project we are classifying links and URL's as malicious or safe using Lexical analysis of strings and machine learning classifier. we have also used features like page rank and DNS features to check the authenticity of that link , we used logistic regression algorithm for prediction because it is giving High accuracy . we check the parameters in link for classification like [no of @, no of /, no of __] and using all these features we are able to classify the link as safe or unsafe

ROLE: **1)**Created a front end using HTML, CSS and Bootstrap **2)** preprocess the data (remove duplicated values and remove Nan) **3)** created machine learning models Like SVM and Logistic regression

3.Text classification using BERT (internship Project): In this project we are trying to classify the text , in previous methods like LSTM and RNN we are not able to understand the context (meaning) of text hence sometime it is hard for classification . But in this project we are using Transformers based NLP state of the art models which is able to understand context from both direction of sentence . So it is robust in classification , I'm using pytorch and BERT-small uncased model for model making and performing EDA using sklearn library this is project is under construction

ROLE 1) gathered Data from Kaggle 2) text preprocessing (stopword removal) 3) use skearn to perform train test split 4) use pretrained BERT TOKENIZER to tokenize the text 5) Use **[CLS] ,[SEP] [MASK]**tokens for data processing .creating Encoding od the data and covert text features to encoding 5) choose the max length and sentence . sentences which are small than max length are padded with zero in the end and sentence bigger than max length are Chop off 5) creating a pytorch dataLoder class to load the data

Recommendation Engine (ongoing internship project): created a recommendation Engine using user - user similarity which can be used for similar product recommendation In this project I used , content based , user-user and Item-item similarity approach , I used Matrix factorization And Singular Value decomposition for Dimensionality reduction

ROLE: this model can be used for different-different use cases with some tuning(for similar product suggestion and songs of movie recommendation) 1) gathered the data 2) done data preprocessing calculate text similarity 3) use cosine similarity and pairwise distance to calculate text similarity 4) remove duplicates text 5) removed outliers 6) **convert high amount of data to low amount by converting 2D array to sparse array using SciPy (32 GB to approx 1.5 GB** 7) convert text to vectors using TF-IDF and Word2Vec and calculate similarity between then 8) used **KNN** for choosing same products

Education

Certificate	Board/University	Year
B.E -Computer	PUNE University	2017-2021
HSC	Maharashtra Board	2015-2017
S.S.C	Maharashtra Board	2015

PERSONAL DETAILS

Name **kartik kiran karande**

Date of Birth **2nd Sep 1999**

Languages **Marathi, English ,Hindi ,German**

Passport **applied**

Phone No **9527722700 / 7972852912**

E-mail id karandekartik777@gmail.com

Place: Pune

Date: 27/08/2021 **(kartik karande)**