Abstract Template

Abstract Title: From an archived data field to a world class product feature for customer-experience using NLP and machine learning

OR

Precise Location Finder using NLP and machine learning

Divya Choudhary, Data Scientist

Abstract (not more than 300 words)

DESCRIPTION:

Who would have imagined that a random chat message or note written in a local language sent by customers to their drivers while waiting for a ride/car to arrive for their pickup can be utilised to carve out unparalleled information about pickup points, their names that sometimes even Google map has no idea of & to finally help in creating a world class customer pick-up experience feature!

This talk will focus on details of how **machine learning & natural language processing** can be used on customer text data along with their bookings data to come up with a product feature enabling customers to see all pickup gates near him/her with their appropriate names that they can understand & choose from while booking a car or ride service.

Like any other service company, customer experience while booking a service is of prime importance at GOJEK, a technology startup based in Jakarta, Indonesia that specialises in ride/car hailing along with 17+ other services. With immense data influx in the system from more than 18 services, the data field that have been archived earlier turned out to be the best data that improved how customers book rides/car on GOJEK app. If you want to understand through examples, the untapped value your data can add to your product through curated machine learning and natural language processing, no use case can better explain it than this.

Divya Choudhary, the speaker will speak in detail about this use case- problem statement, solution outline, data processing, major algorithmic decisions/learnings & final output feature. The presentation will highlight on how the product feature was built, focussing majorly on two algorithmic piece that constituted the problem solution:

1. Machine learning clustering technique
    1. DBSCAN vs K-means, how to know when to use what
2. Wonders of Language modeling
    1. Pre-processing of corpus is the key
    2. Great potential of N-gram modeling

What's the takeaway for the audience?

- Understanding of driving any big product feature through data science and machine learning
- Understanding greatness, know-how of n-gram language model
- Knowledge of DBSCAN & k-means clustering
- Motivation to generate wonder from a plethora of unexplored data fields
-

Keywords—Natural Language Processing, Language Modeling, Clustering, Machine Learning, Data Science products, N-gram language model

Professional Biography (100-150 words)

Divya Choudhary is actually a computer science engineer from BIT Mesra, deemed university in India. She completed her major in 2014 and chose the field of data science as profession. She is currently working as a Data Scientist with a Jakarta based technology startup named GO-JEK. She is responsible for building algorithms and mathematical models to drive features across diversified products at GO-JEK.

With 4 years of work experience, Divya is a computer science engineer who has traversed her professional career from being an analyst to a decision scientist to a data scientist. The crux to any data science solution lies in having a problem-solving mindset & Divya has been known for her business acumen & problem solving approach across all the start-ups that she has been a part of.

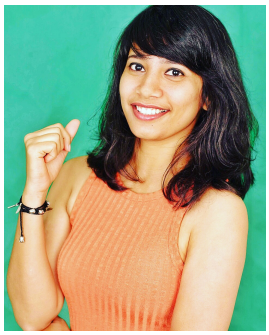
Recent Publications (up to 5)
Recent Major Speaking Experiences:

1. Building Language model from scratch in Python - PyCon 2018 Hyderabad, India (6th Oct 2018)
https://in.pycon.org/2018/#schedule

2. Creating Automated Daily Run Sheet - The Fifth Elephant 2018, Bangalore, India - 24th Nov 2018
https://fifthelephant.in/2018-logisticsconf/
Here is the video of the talk : https://hasgeek.tv/fifthelephant/logistics-in-mobility-conf-2018/1609-daily-run-sheet-is-not-daily-manual-work-anymore

3. Finding Precise Location using NLP - Hack Session - Data Hack Summit 2018 by Analytics Vidhya, Bangalore, India - 22nd Nov 2018
https://www.analyticsvidhya.com/datahack-summit-2018/speakers/divya-choudhary/

4. Accepted for Strata conference San Francisco



Email Id: chowdharydivya28@gmail.com

Presenting author details
Full name: Divya Choudhary
Contact number: +91-7760077955
Twitter account: divya_chy (https://twitter.com/divya_chy)

Linked In account: https://www.linkedin.com/in/divyachoudhary28/
Session name/ number:
Category: (Oral presentation/ Poster presentation) : oral presentation

Any Comments: