

Introduction to Text Analytics for Social Media Chatter

Nabanita Roy

Data Scientist @ ACI Worldwide



PyCon SK 2022

What is Natural Language Processing?



- Process information in the form of natural language or texts
- Combination of Linguistics, Cognitive Science, Statistics
- Enables machines to interact with humans like humans!

HI!
BONJOUR
HOLAI



PyCon SK 2022

Why Social Media Analysis?

Volume
velocity
variety

- A plethora of data, increasing every second
- Individuals and organizations - conversations contain usable information
- It is a two-way street! Bi-directional communication is facilitated by Social Media



PyCon SK 2022

Follow

Types of Social Media Platforms

*not limited to...



Social networks



Blogs / Microblogs



Discussion Forums



Social bookmarks

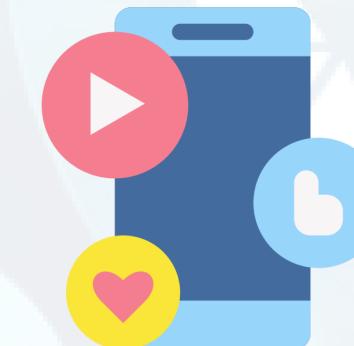
Wiki Projects



Social News



Media Sharing



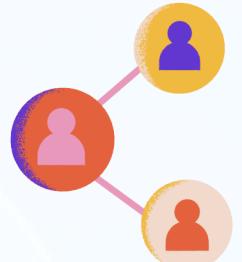
PyCon SK 2022

NLP Tasks for Social Media

*not limited to



- Classification and Clustering
 - Behavior analysis
 - Event detection
 - Customer Segmentation
 - User Modeling
 - Topic Modeling
 - Trends
- Opinion Mining
 - Sentiment, Emotion
 - Sarcasm
- Relationship Extraction
- Named Entity Recognition
- Language Identification and Translation
- Text Summarization
- Content filtering
 - Hate speech, violence, harassment
 - Spams and scams
 - Inappropriate content
 - Rumour Detection



Application of Social Media Analysis

**not limited to*



- All Industries
 - Business Intelligence
 - Recruitment
- Healthcare
 - Mental Health
 - Suicide Prevention
- Defense and security
- Media, Journalism, and Social News
- Marketing
- Finance
- Politics
- Environmental, Social & Governance (ESG)
 - Disaster Management and Recovery
 - Policy Analysis



PyCon SK 2022





NLP is Hard

- Segmentation
- Idioms and phrases
 - *Happiness is a piece of cake*
 - *I called her and gave her a piece of my mind*
- Synonyms
 - *Ecstasy is a piece of cake.*
- Homonyms
 - *A haunting piece of music.*
- Homophones
 - *Happiness is a peace of cake.*
- World Knowledge
- Multiple languages with diverse syntax
- Contractions and more...



Traditionally, NLP techniques assumed the texts are:

- Grammatically correct
- Standard syntax
- Monolingual
- Correct Spellings
- Used formal language



NLP for Social Media is Harder



PyCon SK 2022

Social Media Data

Where to find it?



PyCon SK 2022

Social Media Data

NLP Techniques to Process it



PyCon SK 2022

Social Media Data

NLP Techniques to Process it



PyCon SK 2022