greatlearning
Learning for Life

MACHINE LEARNING- WEEK-3

greatlearning
Learning for Life

DSBA CURRICULUM DESIGN

FOUNDATIONS

- Data Science Using Python
- Statistical Methods for Decision Making

CORE COURSES

- Advanced Statistics
- Data Mining
- Predictive Modelling
- Machine Learning(Week-3/5)
- Time Series Forecasting
- Data Visualization
- SQL

DOMAIN APPLICATIONS

- Financial Risk Analytics
- Marketing Retail Analytics

greatlearning
Learning for Life

LEARNING OBJECTIVE OF THIS MODULE

- Supervised Learning : KNN & Naïve Bayes
- Ensemble Techniques: Bagging, Boosting, Cross-validation and SMOTE
- Text Mining & Sentiment Analysis

greatlearning
Learning for Life

LEARNING OBJECTIVES OF THIS SESSION

- Text Mining
- Sentiment Analysis
- Classification

greatlearning
Learning for Life

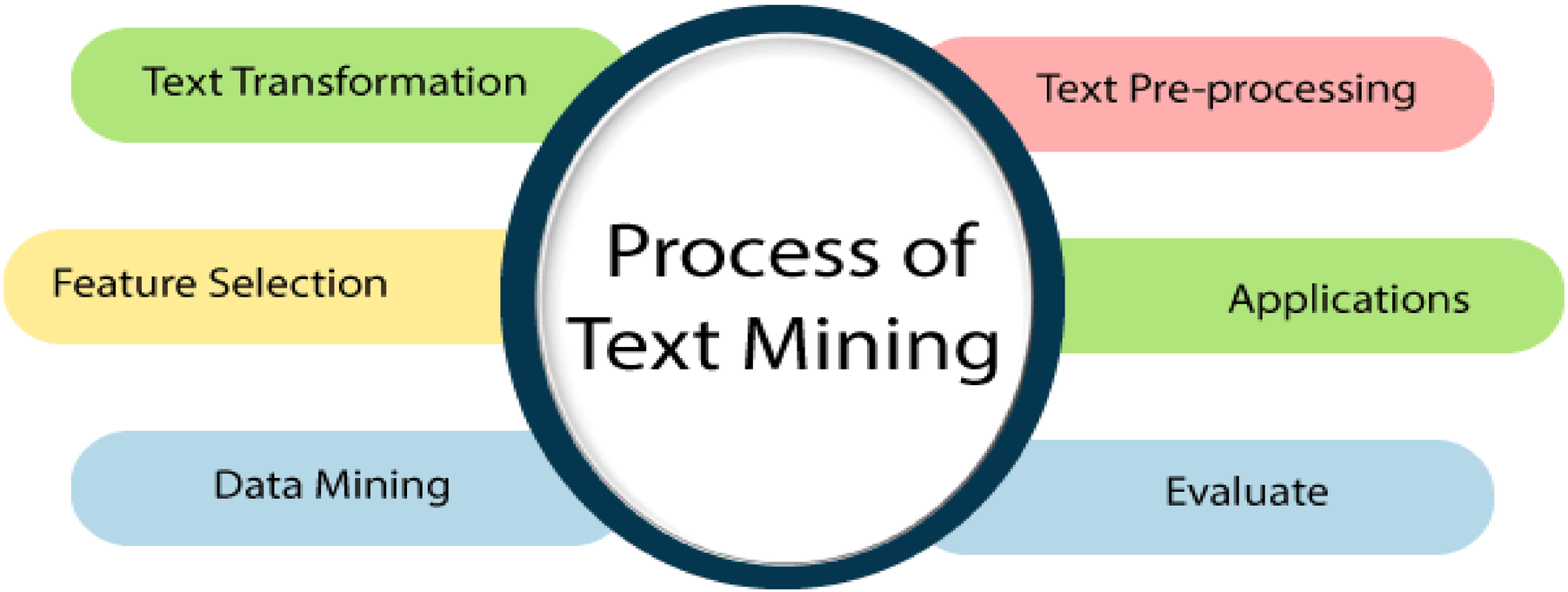
TRY ANSWERING THE FOLLOWING

- What is a document text matrix(DTM)?

- Which is the difference between Unigram and Bigram?

greatlearning
Learning for Life

BROAD OVERVIEW



greatlearning
Learning for Life

BROAD OVERVIEW

greatlearning
Learning for Life

Industry Application - Web scraping: Before making a purchase decision

Often the data available on web pages has a lot of qualitative information with limited instances of quantitative information. With the help of web scraping you can separate the qualitative and quantitative information because the nodes present on a webpage help you organize your data in a much familiar tabular format. You can extract data from multiple sites to compare the prices of the same product on multiple sites.

Likewise it is next to impossible to read through all the reviews of a product before you make a purchase decision, but if you want to have a quick overview of what the existing buyers of that product have experienced, you can extract just the feedback node i.e. where the customers' reviews are present. Now you can create a word-cloud or perform sentiment analysis in order to make a more informed decision.

Reference: https://towardsdatascience.com/https-medium-com-hiren787-patel-web-scraping-applications-a6f370d316f4

Industry Application - Why it matters to show up on top in a search result?

Ever thought of how the results are displayed when you trigger a google search? Out of billions of web pages how is it that google displays the most relevant ones for you? Note: We are referring to organic search results and not the paid ads.

There are web crawlers or you can say bots which filter out the websites that contain the content most relevant to your search. They also have a ranking mechanism which shows you the most frequently visited and engaging web pages on top. Which means you get the results not just in random order but there is a lot of organized thought(algorithm) that goes in the background.

Today, more than the physical presence the businesses are concentrating on web presence. You would agree that it is rare that we click on the second page of the search results often we find what we are looking for on the first page itself. Therefore, the enterprises are willing to spend a good fortune to just show up on the first page of the search results. This is obtained using a number of practices which lead to SEO or Search Engine Optimization.

Reference: https://vnju.org/2018/10/25/search-engine-optimization-seo-strategies-using-text-mining/

greatlearning
Learning for Life

CASE STUDY- Apple Tweets

Using data as a data is a difficult task, as text data is not structured as according to the requirement and not well written, use of the symbol and other symbolic representation make text analytics more difficult. so handling text data is a challenging problem. So, for this field is called Natural Language Processing comes, goal of NLP is to understand and derive meaning from human language in a meaning full way so that machine can understand.

We will be trying to understand sentiment of tweets about the company Apple, By using the twitter for better understand public perception, Apple wants to monitor how people feel over time and how people receive new announcements. Our challenge is to see if we can correctly classify tweets as being negative, positive, or neither about Apple.

Data Science @ Work

greatlearning
Power Ahead

Apply **Data Science at your workplace** to gain some instant benefits:

- Get noticed by your management with your outstanding analysis backed by data science.
- Create an impact in your organization by taking up small projects/initiatives to solve critical issues using data science.
- Network with members from the data science vertical of your organization and seek opportunities to contribute in small projects.
- Share your success stories with us and the world to position yourself as a subject matter expert in data science.

greatlearning
Learning for Life

ANY QUESTIONS

greatlearning
Learning for Life

HAPPY LEARNING

Excelerate CareerPrep:
Career services by Great Learning.

The CareerPrep is an organized way to learn about all of the topics asked in interviews for Data Science jobs.

- Complete Excelerate career prep and learn how to crack an interview with ease.
- Understand the different kind of roles available in the data science ecosystem along with their skill set requirements
- Build a resume that has a higher probability of getting shortlisted by recruiters
- Prepare for different data science roles by reviewing 400+ questions asked in actual interviews
- Understand and implement frameworks to solve guesstimates, logical and aptitude problems often used in interviews
- Learn from the experience of alumni who have successfully transitioned
- Speed up your path towards a transition to a Data science role.

Complete Career Prep today to speed up your path towards DS transition. You'll have a 40% better chance of getting hired than other candidates.

Note: CareerPrep is mandatory step in unlocking the job board.