

## Book Review

### **Test Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from Your Data**

Dipanjan Sarkar, Apress®California (2016), Springer Science Delaware.

ISBN:- 13(pbk):978-1-4842-2387-1

DOI:- 10.1007/978-1-4842-2388-8

*Shabir A. Bhat\**

This book is based on natural language processing, using large SQL databases for text analysis using a programming language called Python which is form of artificial intelligence. The main challenges for analyzing the texts is trying to extract meaningful patterns and useful insights that would be beneficial to the organizations and to store the textual data in Hadoop Distribution File System and access it as needed, which is one of the main principles of a data lake. The book follows a comprehensive and structured approach and is divided into seven chapters.

First, it tackles the basics of natural language understanding and Python constructs in the initial chapters. Once the basics are understood, it addresses interesting problems in text analytics in each of the remaining chapters, including text classification, clustering, similarity, analysis, text summarization, and topic models. In this book, text structure, semantics, sentiments, and opinion are analyzed. The author has used data from tweets and comments on the social media websites. For each topic he has covered the basic concepts and used some real-world scenarios and data to implement techniques covering each concept. The idea of the book is to give a flavor of the vast landscape of text analytics and natural language processing and arm the learners with the necessary tools, techniques and knowledge to tackle our own research problems and start solving them.

We have ushered in the age of Big Data where organizations and business are having difficulty managing all the data generated by various systems, processes and transactions. However, the term ‘Big Data’ is misused and a lot due to the nature of its popular but vague definition of “the 3 V’s”- volume, variety, and velocity of data. This is because it is sometimes very difficult to quantify exactly what ‘Data’ is Big. Some might think a billion records in a database would be Big Data, but that number really seems minute compared to the *petabytes* of data being generated by various sensors or even social media. There is a large volume of unstructured textual data present across all organizations, irrespective of their domain. Just to take some examples, we have vast amounts of data in the form of tweets, status updates, comments, hashtags, articles, blogs, wikies and much more on social media. Even retail and e-Commerce generate a lot of textual data from new product information and metadata with customer reviews and feedback.

There are standard text normalization procedures or a classifiers, and the author has addressed expanding contradictions, text standardization through lemmatization, removing special characters and symbols, and removing stop words in the book, but not focused too much on correcting spelling and

\*Professor, Department of Management Studies, University of Kashmir.

other advanced techniques. Also, the author has addressed the necessary code files along with the detailed explanation of each function.

The author has analyzed various lexical semantic relations from the first chapter using sunsets and real-world example and looked at relationships including entailments, homonyms and homographs, synonyms and antonyms, hyponyms and hypernyms, and homonyms and meronyms. Semantic relations and similarity computation techniques have also been discussed in detail with examples that leveraged common hypernyms among various sunsets. Some popular techniques widely used in semantic and information extraction including word sense disambiguation and named entity recognition with suitable example have also been discussed. Besides semantic relations the concepts related to semantic representations, namely propositional logic and first order logic have also been incorporated.

Next, the author discusses the concept of sentiment analysis and opinion mining and their domains of social media, surveys and feedback data. He has taken practical example on analyzing sentiment on actual movie reviews from IMDb and built several models that include supervising machine learning and unsupervised lexicon-based models which brings and end to the book.

My critical evaluation of the book is that it consist of algorithms used in Python to analyze the textual data which is a bit difficult for the people to understand who does not have at least working knowledge of the algorithms or the basic programming languages like C, VB and object-oriented programing. The syntax of the code has to be understood and remembered, and then only the objectives of the text analytics could be achieved which in a drawback and less familiarity of the people with Artificial Intelligence. It will be irrational to say that the text analytics are easier in Management as a discipline but keeping in view the dynamic nature of research the new methods have to be diffused and adopted.

The book provides a clear understanding of a language developed and evolved with by human through natural use of communication rather than constructed and created artificially, like a computer programming language. The book is therefore a significant contribution to the existing literature and recommended for use in text analytics.