BOOK REVIEW

TITLE OF THE BOOK: DATA SCIENCE IN MARKETING ANALYSIS

AUTHOR: MIRZA RAHIM BAIG, GURURAJAN GOVINDAN, And VISHWESH RAVI SHRIMALI

PUBLISHER: PACKT PUBLISHING

YEAR: 2021

Reviewer: Hari Thapliyal,
AI Educationist & Consultant, Founder, dasarpAI

INTRODUCTION

Before writing this book, three authors of this books has written 3 books separately. Mirza Rahim Baig has authored "The Deep Learning Workshop", Gururajan Govindan authored "The Data Analysis Workshop" and Vishwesh Ravi Shrimali has written a book "The Computer Vision Workshop". The individual's experience of writing book on Data Science has made this book easy understandable and lucid book. I am doing this review on the second edition of this book. The first edition was published in 2019 and the second edition is published in 2021. This book contains 637 pages, and it looks heavy on your table but most the content is code example and screenshots of graphs and data tables. This is normal for any good programming book, which wants their reader to understand what code was written and how the output looks like. This book is written in simple English. In the book, authors trying to explain how to work with python, pandas, seaborn and other important python libraries. This book goes beyond the title of this book. It is not only about marketing analysis but also how to develop model using scikit-learn and other machine learning libraries. Copyright of this book is with Packt Publishing.

EVALUATION

The content of this book is more than the name of this book. It is not book only for marketing analysis but goes beyond the analysis. In this book, authors discuss important steps of data cleaning, data exploration, data visualization. After that the books discusses unsupervised learning four important algorithms Kmean, Kmode, mean-shift and K-prototypes. We know in any k based algorithm the first challenge is to guess the correct k. Authors discusses elbow method and silhouette score to determines the k. After this, the book proceeds

towards developing regression. Authors has explained different metrics to evaluate the performance of linear-regression based model. Tree based regressions and random forest are also explain for developing regression model. Another supervised learning type is classification. Algorithms, cost function, metrics required to develop, evaluate classification models are discussed very lucidly. Some popular classification algorithms like support vector machine, decision tree, random forest are discussed along with the hyperparameters tuning. Authors discusses four performance metrics of classification models namely precision, recall, f1 score and roc curve. Sometimes target variable of our datasets is not balanced and it is not straight forward to train classification model using this kind of dataset. This kind of dataset is called imbalanced dataset. Techniques to handle class imbalance is also discussed in this book.

This books has nine chapters. All the code written to explain the working is written in python. All what you need to execute the code of this book is pandas, matplotlib, seaborn, numpy and scikit-learn library along with python3. The books starts from the basic of installing python, conda and creating virtual environment using conda and install python packages in the conda environment. This book

has many exercises and activities with every chapter. If you spend time and practice along, with the given code then you will enjoy reading this books and problem solving in data science.

This book is not heavy or complex by any standard. Examples are good and code is simple. The code is available in the github repository at https://packt.link/59F3X

Chapter 1: Data Preparation and Cleaning. Preparing a clean dataset using panda's library. Data cleaning, missing data handling, joining data files, grouping data, inspecting data structure.

Chapter 2: Data Exploration and Visualization. Use of groupby, unique, value_counts, shape, pivot functions. Visualizing dataset with Seaborn and Matplotlib.

Chapter 3: Unsupervised Learning and Customer Segmentation. Working of k-mean clustering, calculating distance between vectors, scaling data is discussed in this chapter.

Chapter 4: Evaluating and Choosing the Best Segmentation Approach. Techniques to evaluate quality of cluster, number of cluster and 4 types of clustering methods are discussed.

Chapter 5: Predicting Customer Revenue Using Linear Regression. Feature engineering for regression model, creating linear regression model and interpreting this model.

Chapter 6: More Tools and Techniques for Evaluating Regression Models. Recursive feature selection, RSE, MAE, Decision Tree and Random forest, these techniques and algorithms helps you avoiding in creating an over fitted model.

Chapter 7: Supervised Learning: Predicting Customer Churn. Defining classification problem, understanding logistic regression, cost function of logistic regression, building classification model and interpreting this model.

Chapter 8: Fine-Tuning Classification Algorithms. Making robust model using SVM, hyper parameter tuning for DT and RF based model. Scaling feature using normalization and standardization. Techniques to evaluate classification model.

Chapter 9: Multiclass Classification Algorithms. Dealing with class imbalanced dataset.

CONCLUSION

This book is for following professionals.

A. Those who have some industry experience and can relate to the business problems.

B. Those who like learning new programming language and want to start their journey into machine learning or analytics.

C. Those want to switch their career from typical software development to data science.

D. Those who love visualizing and summarizing the data.

Using practical business cases authors have tried to teach how to clean the data, analyses the data, create machine learning model, and evaluate the performance of those models. Overall, this book serves the purpose for which it was written.