CRANFIELD UNIVERSITY

CRANFIELD SCHOOL OF MANAGEMENT
MSC LOGISTICS AND SUPPLY CHAIN MANAGEMENT
MSC PROCUREMENT AND SUPPLY CHAIN MANAGEMENT

BIG DATA ANALYTICS

ASSIGNMENT

Date Set: 4 March 2020
Date Due: 9 April 2020

Introduction

In this Big Data Analytics Assignment, you will be assessed on two areas: theory and practice. In the theory area, we expect you to critique recent publications in the big data analytics domain. In the practice area we present you with a dataset on which we expect you to perform appropriate analysis and draw managerial conclusions. Each of these areas will comprise 45% of your marks. As is the case in many other assignments, 10% of the marks is attributable to style and presentation.

The word limit for this assignment is 1,500. It is an upper limit, not a target. Please report the number of words you use and do not exceed this limit. Although theory and practice questions are equally weighed, we anticipate you will use more words in the theory part, roughly 2/3 of your upper limit.

Please use the discussion board for any questions you might have about this assignment. We will answer all questions that are asked on or before 6 April 2020 12pm to minimise the last-minute stress and encourage timely attention to the assignment.

It has been an absolute pleasure to teach you big data analytics. We hope you enjoy this assignment and use the techniques in the future.

Abhijeet Ghadge, Emel Aktas, and Qiongyu Mastropierro

March 2020, Cranfield

Disclaimer: Although the problems presented in this assignment are informed by real events, the names, characters, businesses, places, events, locales, and incidents are the products of the authors' imagination. Any resemblance to actual persons, actual businesses, or actual events is coincidental.

Q1 Theory [45 marks] **transport route optimisation**

Using the supply chain operations reference model, identify an area within which you will search for academic articles explaining the use of

1. Data mining, or
2. Neural networks, or
3. Probabilistic graphical models

Using your favourite research database (e.g. Scopus) search and select three papers that explain the application of any of the above methods to solve a business problem within any area of the supply chain operations reference model. Critique these papers in terms of their aims and objectives, data sources, methods and findings, and key conclusions. Reflect on whether you argue with the main thesis and propose conditions where the proposed model may fail.

Don't forget to attach the three papers you have identified as an appendix.

Q2 Practice [45 marks]

For the practical example, you are given a data set of products in which the company wishes to determine which products it should continue to sell, and which products to remove from their inventory. The file contains historical sales data and active inventory, which can be discerned with the column titled "File Type".

We expect you to choose a method and explain why you chose that method. Decision Tree, Logistics Regression, Neural Networks are possible approaches for this analysis.

If you applied big data analytics methods on this dataset (e.g. decision tree, logistic regression, or some other machine learning model) can help the company generate a value (i.e., probability score) for each product, that can be used as the main determinant when evaluating the inventory. Each row in the file represents one product. There are many products in this dataset and few of them tend to sell (only about 10% sell each year) and many of the products are sold only once in a year.

The file contains historical sales data (identified with the column titled File_Type) along with current active inventory that should be evaluated (i.e., File Type = "Active"). The historical data shows sales for the past 6 months. The binary target (1 = sale, 0 = no sale in past six months) is likely the primary target that should drive your analysis. Other columns contain numeric and categorical attributes that are considered relevant to sales.

When you analyse this data, you will observe that some of the SKUs with historical sales are also included in the active inventory. The company keeps a record of the following attributes, but not all of them would be relevant to your analysis.

1. **Order:** A sequential counter. No further information captured in this.
2. **File_type:** If historical, the information applies to past six months, if active, it is currently in the inventory waiting to be sold.
3. **SKU_number:** A unique identifier for each product.
4. **SoldFlag:** Equals to 1 if the SKU is sold in past 6 mos, 0 otherwise.
5. **MarketingType:** Two categories of how the product is marketed. This can probably be ignored, or, each type can be considered independently.
6. **ReleaseNumber:** The counter for the number of releases the product had, 0 for new product launch.
7. **New_Release_Flag:** Any product that has had a future release (i.e., Release Number > 1)
8. **StrengthFactor:** An estimate of the market size of the product, reliability of this estimate is questionable. **test/ correlation**
9. **PriceReg:** Regular price
10. ReleaseYear: The year in which the product was released.
11. ItemCount: Number of items in stock
12. **DiscountedPrice:** Price when the product goes on discount.
13. **PromotionPrice:** Price when there is a promotion on the product (bundle, rather than discount, though there is no data on with what it was bundled).

Please develop a model that will provide this company with a probability estimate of a sale for each of their SKU. Please provide an evaluation of the accuracy of your selected model.

Style and Presentation [10 marks]

You will receive marks for the style and presentation of your assignment. Please pay attention to

1. Writing grammatically correctly and concisely.
2. Providing captions for your tables and figures and citing them in the text.
3. Using appropriate level of accuracy (no need for 5+ figures after the decimal point. We recommend 3).
4. Presenting your assignment in a report structure so we can map your answers to the questions we have asked.
5. Citing the references you used in preparing your answers.

Marking Criteria

Marks will be awarded for the **level of understanding demonstrated in the concepts** and **accuracy of your answers**.

explain the reason why like it/ not like it

ASSESSMENT CRITERIA	**Excellent (80-100%)**	**Very good (70-79%)**	**Good (60-69%)**	**Satisfactory (50-59%)**	**Poor (40-49%)**	**Very poor (0-39%)**
Theory [45] • Publication time of papers [5] • Relevance of papers [10] • Comparison [20] • Reflection [10]	☐Papers selected are recent. ☐ Papers are relevant to supply chain management. ☐ Appropriate comparison criteria are identified and applied competently. ☐ Reflections are detailed and referenced.	☐Papers selected are published within the last five years. ☐ Papers are mostly relevant to supply chain management. ☐ Mostly appropriate comparison criteria are identified and applied mostly competently. ☐ Reflections are detailed.	☐Papers selected are published within the last decade. ☐ Papers are somewhat relevant to supply chain management. ☐ Somewhat appropriate comparison criteria are identified and applied somewhat competently. ☐ Reflections lack sufficient detail at times.	☐Papers selected were published more than a decade ago. ☐ Papers have questionable relevance to for supply chain management. ☐ Comparison criteria identified have questionable appropriateness to the task at hand and evaluation of the papers is questionable. ☐ Reflections are general and basic.	☐ Papers selected were published more than a decade ago. ☐ Papers are not relevant to supply chain management. ☐ Comparison arguments are not valid at times. ☐ Reflections are not specific to the area identified.	☐ Papers missing. ☐ Papers missing. ☐ Papers not compared. ☐ There are no reflections.

executive recommendation (eg. slow-moving items, how to deal with)

ASSESSMENT CRITERIA	Excellent (80-100%)	Very good (70-79%)	Good (60-69%)	Satisfactory (50-59%)	Poor (40-49%)	Very poor (0-39%)
Practice [45] • Descriptive Statistics [10] • Model Build and Test [25] • Insights [10]	☐ Descriptive statistics for each variable is produced and presented clearly with conclusions. ☐ The steps to build the model explained well and the model choice justified. ☐ Data-driven, meaningful recommendations are produced for the company and justified by references and relevant examples.	☐ Descriptive statistics for many of the variables is produced and presented clearly with key conclusions. ☐ The steps to build the model explained without any justification. ☐ Several meaningful recommendations produced for the company and justified by relevant examples.	☐ Descriptive statistics for few of the variables is produced and presented with some. ☐ The main steps to build the model are explained. ☐ Few recommendations are produced for the company and justified by somewhat relevant examples.	☐ Descriptive statistics for key variables is produced and presented without any commentary. ☐ Main steps to build the model are presented without any explanation. ☐ Few recommendations are produced for the company without justification.	☐ Descriptive statistics produced has some errors. ☐ Model build steps and choices made by the author not explained. ☐ The link between recommendations and the results is missing.	☐ Descriptive Statistics missing. ☐ No model built. ☐ No recommendations produced.

ASSESSMENT CRITERIA	Excellent (80-100%)	Very good (70-79%)	Good (60-69%)	Satisfactory (50-59%)	Poor (40-49%)	Very poor (0-39%)
Style and Presentation [10] • Coherence and conciseness of writing. Good grammar, no spelling errors, and use of appropriate vocabulary. • Relevant and accurate referencing • Effective and well-presented figures and tables	☐Exceptionally fluent structure, and clarity of expression. ☐ Referencing clear, relevant and consistent. Virtually error-free. ☐ Table and figure captions are meaningful. Tables and Figures cited in the text. No spillage on the page margins.	☐ Language fluent thoughts and ideas clearly expressed. ☐ Referencing, relevant with few errors. ☐ Table and figure captions are meaningful. Tables and Figures are consistently cited in the text.	☐ Language mainly fluent, minor spelling and/or grammar and/or punctuation errors. ☐ Referencing, relevant but with several errors. ☐ Table and figure captions are sometimes meaningful. Tables and Figures are sometimes cited in the text.	☐ Language understandable, meaning apparent but not explicit, grammar and/or spelling poor. ☐ Inconsistent, referencing with some errors. ☐ Table and figure captions with minimal information. Tables and Figures are not always cited in the text and sometimes spilling over the margins.	☐ Language far from fluent but understandable, grammar and/or spelling poor. ☐ Inconsistent, incoherent referencing. ☐ Table and figure captions missing or inconsistent. Tables and Figures cited wrongly in the text.	☐ Language far from fluent, meaning unclear, grammar and/or spelling poor. ☐ Incoherent and/or absent referencing. ☐ Table and figure captions are missing. Tables and Figures not cited in the text.