THE UNIVERSITY OF HONG KONG

MSc in E-Commerce and Internet Computing

ECOM7126 Machine Learning for Business and E-Commerce (2023-24)

Group Mini-Project

The Group Project for ECOM7126 consists of 2 parts: (1) a Case Study and (2) an End-to-End Machine Learning Project in a business situation. Each part carries equal marks of 50 each for the Mini-Project. The Group Mini-Project counts for 40% of the overall course assessment.

Your report should include a brief summary of the contribution by each member of the team (i.e.who does what), AND there will be a confidential Peer Evaluation conducted at the end of the Group Project. For details, please refer the Moodle course website. The use of AI and GPT-related tools follows that of the University guideline. Use of GPT to generate reports or program codes are strictly prohibited. There may be an optional oral check on whether any declared work by a member of the team is actually done by that member.

Deadline: 14 May 2024

Part 1 - Case Study (50 marks):

How Machine Learning Will Transform Supply Chain Management
Narendra Agrawal, Morris A. Cohen, Rohan Deshpande, and Vinayak Deshpande
Harvard Business Review, Magazine (Mar – Apr) 2024

Discuss among members in your group (every member should read the entire Article and contribute). Write a short paper summarizing your observations and takeaways from this article, include but not limited to the following questions:

(1) What are the challenges in Supply Chain that the paper want to address, and with what objectives?
(2) In what way would ML transform supply chain management?
(3) Discuss and critique the authors' claim that their approach does a better job in managing supply chain.

Your paper should be limited to about 3,000 words (soft limit), and should cover, but not limited to, the above three questions. All students in the group should contribute.

(Note: The article is made available to students of ECOM7126 for personal study ONLY under a paid license to Harvard Business Review. Copy or distribute any copy of this Case Study Report is strictly forbidden, otherwise you personally would have infringed the Copyrights Law).

Part 2 – The Orchard, Singapore (50 marks):

The Orchard is a renowned, luxury hotel chain in Singapore. The hotel chain has one business hotel (CT) close to the business district, near the Orchard Road, and one resort hotel (RS) on the famous Sentosa Island. The hotel had collected 3 years anonymized data about booking and occupancy of these two hotels in order to help managing future booking, inform pricing policy, assist marketing and other business decisions. The General Manager considers one very important challenge (among others) is to estimate cancellation more accurately among other potential insights that may be obtained from these data to ensure as high occupancy as possible.

The following are the feature label descriptions of the data collected for these two hotels:

Feature label	**Description**
hotel	"CT" = Orchard City; "RS" = Orchard Resort
is_cancelled	0 = not cancelled; 1= cancelled
lead_time	# of days between booking and arrival
av_daily_rev	Average Daily Revenue from the booking in S$
arrival_date_year	arrival year
arrival_date_month	arrival month
arrival_date_week_number	arrival week in the year
arrival_date_day_of_month	arrival day of the month
stays_in_weekend_nights	# of weekend nights
stays_in_week_nights	# of weeknight stay
adults	# of adult guests
children	# of children guests
babies	# of baby guests
is_repeated_guest	0 = not repeated guest; 1 = repeated guest
previous_cancellations	# of previous cancellations
previous_bookings_not_cancelled	# of previous bookings not cancelled
room_type	Type of room booked
booking_changes	# of booking changes made
deposit_type	"No Deposit"; "Non Refund"; "Refundable"
require_pickup	# of pickup from airport requested

Your mission is :

1. Analyse the data systematically to get insights.
2. Perform necessary data cleansing and get the data ready for ML modelling.
3. Use at least 2 and not more than 3 different ML models to try to predict cancellation of bookings, and find the best performing one for the management so as to inform sound management decisions.
4. Consider what possible any insights you may have on the business in these two hotels from the dataset if any, and report to the GM.

Your submissions should include a report on your finding with the essential results, plus the complete executable Colab notebook of your ML program. Please DO NOT screen capture your codes, as this would not be executable. Rest assured that your work will be tested for its integrity and consistency with your report. While there is no word limit on your report, you are reminded that quality is more important than the size of the report! You should always comment your program codes so that it can be read by others.

Dataset: Orchard_dataset.csv