

Name:

Enrolment No:



**UNIVERSITY OF PETROLEUM AND ENERGY STUDIES**  
**End Semester Examination, December 2018**

**Course: Predictive Modelling**  
**Programme: MBA Business Analytics**  
**Time: 03 hrs.**  
**Instructions:**

**Semester: III**  
**Course Code: DSBA 8003**  
**Max. Marks: 100**

All Questions are **COMPULSORY**. Each question is of **25 Marks**.

There are the following MS Excel data files on your desktop: **Taxi-cancellation.xlsx**; **Cosmo.xlsx** and **eBayAuctions.xlsx** corresponding Questions 2, 3 and 4 of this question paper respectively. Use the data of each of the excel sheets to answer the respective questions and write the solutions in the answer book provided.

Once you complete the question paper save all your excel sheets in a folder with your **SAPID (5000abcd)** as the folder name.

S. No.		Marks	CO
Q. 1	<p>Assuming that predictive modelling techniques are to be used in the following cases, identify whether the task required is supervised or unsupervised and give reasons for the choice:</p> <ul style="list-style-type: none"><li>a) In an online bookstore, making recommendations to customers concerning additional items to buy based on the buying patterns in prior transactions</li><li>b) Identifying segments of similar customers</li><li>c) Predicting whether a company will go bankrupt based on comparing its financial data to those of similar bankrupt and non-bankrupt firms</li><li>d) Estimating the repair time required for an aircraft based on a trouble ticket</li><li>e) Automated sorting of mail by pin code scanning</li></ul>	25	CO 1 & 2
Q. 2	<p><b>Case Study: Taxi Cancellations (Data file: Taxi-cancellation.xlsx)</b></p> <p>In late 2013, the taxi company <b>Yourcabs.com</b> in Bangalore, was facing a problem with the drivers using their platform-not all drivers were showing up for their scheduled calls. Drivers would cancel their acceptance of a call and, if the cancellation did not occur with adequate notice, the customer would be delayed or even left high and dry.</p> <p>Bangalore is a key tech center in India, and technology was transforming the taxi industry. <b>Yourcabs.com</b> featured an online booking system (though customers could phone in as well) and presented itself as a taxi-booking portal.</p> <p>The Uber ride sharing service would start its Bangalore operations in mid-2014. <b>Yourcabs.com</b> had collected data on its bookings from 2011 to through 2013, to see what it could learn about the problem of cab cancellations. The data presented for this case are a randomly selected subset of the original data, with 10000 rows, one row for each booking. There are 17 input variables, including user (customer) ID, vehicle model, whether the booking was made online or via a mobile app, type of travel, type of booking package, geographic information, and the date and time of the scheduled trip. The target variable of interest is the binary indicator of whether a ride was canceled. The overall cancellation rate is between 7% and 8%.</p>	25=[ 5+ 5+ 10+ 5]	CO 3 & 4

	<p><b>Questions:</b></p> <ul style="list-style-type: none"> <li>a) How can a predictive model based on these data be used by <b>Yourcabs.com</b>?</li> <li>b) Explore, prepare, and transform the data to facilitate predictive modeling.</li> <li>c) Fit several predictive models of your choice. Do they provide information on how the predictor variables relate to cancellations?</li> <li>d) Report the predictive performance of your model in terms of error rates (the confusion matrix). How well does the model perform? Can the model be used in practice?</li> </ul>		
<p>Q. 3</p>	<p><b>Case Study: Predicting Software Reselling Profits (Data File: Cosmo.xlsx)</b></p> <p>Cosmo Software is a software catalog firm that sells games and educational software. It started out as a software manufacturer and then added third-party titles to its offerings. It recently revised its collection of items in a new catalog, which it mailed out to its customers. This mailing yielded 1000 purchases. Based on these data, Cosmo wants to devise a model for predicting the spending amount that a purchasing customer will yield. The file Cosmo.xls contains information on 1000 purchases. The description of the variables are provided in the Description sheet of the excel file.</p> <ul style="list-style-type: none"> <li>a) Explore the spending amount for the categorical variables</li> <li>b) Explore and comment about the relationship between the spending and the other continuous predictors.</li> <li>c) Fit a predictive model for spending. Remember your model should be parsimonious.</li> <li>d) Based on this model, what type of purchaser is most likely to spend a large amount of money?</li> <li>e) Evaluate the predictive accuracy of the model.</li> </ul>	<p>25</p>	<p>CO 3 &amp; 4</p>
<p>Q. 4</p>	<p><b>Case Study: Competitive Auctions on eBay.com (Data File: eBayAuctions.xlsx)</b></p> <p>The file <b>eBayAuctions.xlsx</b> contains information on 1972 auctions that transacted on eBay.com during May-June 2004. The goal is to use these data to build a model that will classify auctions as competitive or non-competitive. A competitive auction is defined as an auction with at least two bids placed on the item auctioned. The data include variables that describe the item (auction category), the seller (his/her eBay rating), and the auction terms that the seller selected (auction duration, opening price, currency, day-of-week of auction close). In addition, we have the price at which the auction closed. The task is to predict whether the auction will be competitive.</p> <ul style="list-style-type: none"> <li>a) Fit a classification tree using all predictors.</li> <li>b) Is the model practical for predicting the outcome of a new auction?</li> <li>c) What can you say about the predictive performance of this model?</li> <li>d) Based on this tree, what can you conclude from these data about the chances of an auction obtaining at least two bids and its relationship to the auction settings set by the seller (duration, opening price, ending day, currency)? What you would you recommend for a seller as the strategy that will most likely lead to a competitive auction?</li> </ul>	<p>25=[7+5+5+8]</p>	<p>CO 3 &amp; 4</p>