請於答案卷上依序作答，並註明作答的題號

試卷共四大題

1. Please describe different theories that support or oppose the argument that "IT (information technologies) can help firms shrink in size (i.e., help decrease firm size)." Please also describe under what condition(s) IT can help decrease firm size, and under what condition(s) IT can help increase firm size. **(20%)**

2. In the era of machine learning (such as deep learning), data is the king. Large volumes of data are needed for model training purposes. To build machine learning models for business use, firms need to effectively collect, organize, store, and use data (e.g., process-, customer-, employee-related data, etc.) in order to harvest the values of machine learning.
   (a) Throughout the lifecycle of data (including data collection, organization, storing, and use) for machine learning, firms face many important managerial issues (decisions) concerning the management of these data. An example of such managerial issues (decisions) is: the firm needs to decide beforehand "what data should and can be collected." Please identify four additional managerial issues (decisions) concerning the management of data for machine learning. Please also explain why each of the additional managerial issues (decisions) that you just identified is important. **(20%)**
   (b) For the example decision given in (a) (i.e., the decision on what data should and can be collected), is a centralized or decentralized decision-making structure more appropriate? For a centralized decision-making, we mean that decisions are directed to the top of the organization (i.e., the executive level); for a decentralized decision-making, decision rights are given to a lower organizational level (e.g., business units). Please provide a detailed explanation to justify your answer. **(10%)**

3. Minho joined a group in a large IT company that develops customized data analytics solutions for their customers. In his first project, his colleagues collected several hundred thousands of data points and developed a model that can predict whether customers will purchase a product in their next session. The developer claimed that the prediction accuracy is 92% using the standard ten-fold cross-validation. Answer the following questions.
   (a) What is the definition of prediction accuracy? In additional to prediction accuracy, what are other useful performance measures that can help us evaluate the prediction performance? **(6%)**
   (b) Is an accuracy of 92% good or bad? How do we obtain a meaningful evaluation of this performance level? **(7%)**
   (c) The developer used the convolutional neural network (CNN) to learn the prediction model. He believes that this is the state-of-the-art approach and no other approach can perform better. Do you agree with him? Why? **(6%)**
   (d) How can we improve the prediction performance if the customer is not satisfied with the performance? **(6%)**

4. The use of Artificial Intelligence (AI) to assign individuals into classes has stimulated concerns. One major issue is about the fairness of an automated classification algorithm. Consider a fictitious case that the Taitai Bank developed a predictive model for predicting future loan default. The higher score indicates that the individual has a higher probability of defaulting a loan. Taitai Bank has a policy of being fair to all its customers. Their main concern is that the risk score should be fair to customers in different occupations. Internal discussions reveal three definitions of fairness. To simplify our discussion, a "positive case" means that the individual will default in the future and

見背面

a "negative case" means that the individual will not default in the future. The three different definitions of fairness are:

(1) The risk score should be well-calibrated across different occupational groups: if the algorithm identifies a set of people as having a probability of z of constituting positive cases, then approximately a z fraction of this set should indeed be positive instances. This conditional should hold when applied separately in each group.

(2) The average score received by people constituting positive instances should be the same in each occupational group.

(3) The average score received by people constituting negative instances should be the same in each occupational group.

Answer the following questions.

(a) Why does definition (1) corresponds to notions of fairness? **(5%)**

(b) Why does definition (2) corresponds to notions of fairness? **(5%)**

(c) Why does definition (3) corresponds to notions of fairness? **(5%)**

(d) Which of the definitions (1), (2), or (3) is more reasonable? Why? **(5%)**

(e) What should Taitai Bank do to ensure that their risk score is fair? **(5%)**

試題隨卷繳回