**Please solve A, B, C and D.**

1. **Clustering**

Please do the followings from **Chapter 8 (Page 563-567: Tan, Steinbach, and Kumar’s book):**

* Problem#16,
* Problem#17,
* Problem# 23

B. **Classification**

Please do the followings from **Chapter 5 (Page 318-322: Tan, Steinbach, and Kumar’s book):**

* Problem#7
* Problem#8,
* Problem# 9 (a) and 9(b)

1. **Stream Mining**

*These following problems are related to*

* Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, and Bhavani Thuraisingham**,** Classification and Novel Class Detection in Concept-Drifting Data Streams under Time Constraints, *IEEE Transactions on Knowledge & Data Engineering (TKDE), 2011*, IEEE Computer Society, June 2011, Vol. 23, No. 6, Page 859-874.
* [Mohammad M. Masud](http://www.informatik.uni-trier.de/%7Eley/pers/hd/m/Masud:Mohammad_M=.html), [Clay Woolam](http://www.informatik.uni-trier.de/%7Eley/pers/hd/w/Woolam:Clay.html), [Jing Gao](http://www.informatik.uni-trier.de/%7Eley/pers/hd/g/Gao:Jing.html), Latifur Khan, [Jiawei Han](http://www.informatik.uni-trier.de/%7Eley/pers/hd/h/Han:Jiawei.html), [Kevin W. Hamlen](http://www.informatik.uni-trier.de/%7Eley/pers/hd/h/Hamlen:Kevin_W=.html), [Nikunj C. Oza](http://www.informatik.uni-trier.de/%7Eley/pers/hd/o/Oza:Nikunj_C=.html): Facing the reality of data stream classification: coping with scarcity of labeled data. [Knowl. Inf. Syst. 33](http://www.informatik.uni-trier.de/%7Eley/db/journals/kais/kais33.html#MasudWGKHHO11)(1): 213-244 (2011)

1. For Novel class detection, we use semi-supervised K-means algorithm. Let us assume we have two existing classes (+ & -; positive and negative). In a particular training chunk we have 1500 instances associated with + class and 500 instances associated with – class. For that chunk how do you apply K-means when K=60? How many clusters can we generate for + class and for – class?
2. Please explain marginal false positive and marginal false negative instances.
3. Concept-evolution occurs as a result of new classes evolving in the stream. When a new class evolves in the data stream, traditional data stream classification techniques cannot identify them, which results in high classification error. Our previous works in data stream classification addressed the concept-evolution problem along with the other two challenges. We applied an ensemble classifier, where each classifier was equipped with a novel class detection module. However, our previous approach ignored the case of recurring or seasonal classes. This special case occurs when a class appears in the stream, then disappears for a long time, and again appears. In this case, our previous approach would consider the class as novel both in its first and second appearance.
4. Please explain how we address this issue, and propose a more realistic novel class detection technique, which remembers a class and identifies it as “not novel" when it appears after a long disappearance.
5. What is the difference between unsupervised clustering and semi-supervised clustering algorithm?
6. **Association Rule Mining**

Please do the followings from **Chapter 6 (Page 406-408: Tan, Steinbach, and Kumar’s book):**

* Problem#6,
* Problem#7,
* Problem# 8
* Problem#9