CSEE5590-0001/490-0003: Big Data Programming

Lesson Plan # 14

Lesson Title: *Apache Spark MLIB*

Lesson Description: *Apache Spark MLib*
1. Clustering
2. Classification
3. Regression
4. Recommendation

Lesson Overview:
MLlib is Apache Spark's scalable machine learning library, with APIs in Java, Scala, Python, and R.

In class exercise:
1. **Classification:**
   Datasets:
   - **https://archive.ics.uci.edu/ml/datasets/Acute+Inflammations**
   - **https://archive.ics.uci.edu/ml/datasets/Adult**

   Algorithms
   a. **Naïve Bayes**
   b. **Decision Tree**
   c. **Random Forest**

2. **Clustering:**
   Datasets:
   - **https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008**

   Algorithm
   a. **KMeans**

3. **Regression:**
   Datasets:
   - **https://archive.ics.uci.edu/ml/datasets/Automobile**

   Algorithm
   a. **Linear Regression**
   b. **Logistic Regression**

Bonus:
1**.** Show confusion matrix for any machine learning algorithm
2. Calculate Precision. Recall, and F1-score
3. Inference on custom data for any algorithm of your own.

Note: it's expected to have submission in only one coding language. Documentation should include Limitations, Conclusion and your observations.

ICP Submission Guidelines:

1. ICP Submission is individual however, it can be completed as a Team during session.
2. If completed, should be presented to TA or Instructor before the completion of the class
3. Submission after the deadline is considered as late submission. (Check the late submission policy in the syllabus)
4. ICP Code with brief explanation should be pushed to GitHub.
5. Submit your screenshots as well to GitHub and documentation. The screenshot should have both the code and the output.
6. Submit a demo video 2-3 min showing your assignment with a voice over explaining your work if you are unable to complete ICP within the deadline due to genuine reason.
7. Provide the video submission link through the GitHub and submission form https://forms.gle/xMAmr3zATrtMG5cX7

Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL: https://catalog.umkc.edu/special-notices/academic-honesty/