# ONE PAGE COURSE OVERVIEW

**Course Title: D209 – Data Mining I**

**Course Description:**  Data Mining 1 expands predictive modeling into nonlinear dimensions, enhancing the capabilities and effectiveness of the data analytics lifecycle. In this course, learners implement supervised models specifically classification and prediction data mining models to unearth relationships among variables that are not apparent with more surface-level techniques. The course provides frameworks for assessing models' sensitivity and specificity.

**Competencies:**
a.  Classification Data Mining Models: Apply observations to appropriate classes and categories using classification models.
b.  Predictive Data Mining Models: Implement prediction data mining models to find hard-to-spot relationships among variables.
**c.**  Data Mining Model Performance: Evaluate model performance for precision, accuracy, and model comparison.

**Performance Assessment (PA):** Task covering the above competencies and involving:

**Task 1:** Create a data mining report using classification and data mining models as follows:
   a.  Answer a research question using either KNN or Naïve Bayes.
   b.  Justify your classification method, assumptions, list packages and libraries chosen for Python or R.
   c.  Prepare your data and provide a copy of your cleaned data set.
   d.  Analyze, summarize, and discuss the implication of your work.
   e.  Do a Panopto video recording of your report.

**Task 2:** Create a data mining report using predictive data models and data mining model performance as follows:
   f.  Answer a research question using a prediction method such as decision trees, random forests, advanced regression (i.e. lasso or ridge regression).
   g.  Justify your prediction method.
   h.  Prepare your data and provide a copy of your cleaned data set.
   i.  Analyze, summarize, and discuss the implication of your work.
   **j.**  Do a Panopto video recording of your report.

**Tools and Techniques:** Python or R.  (Students can choose); KNN, Naïve Bayes, decision trees, random forests, advanced regression.

**Resources:**
   a.  Datacamp
   b.  Cohorts Webinar - holds the 2nd Sunday of every month. Time: 2:00 pm MDT, 3:00 pm CST, or 4:00 pm EST.
       https://wgu.webex.com/wgu/k2/j.php?MTID=tec3de85d33283f764cc2049a269827b4