

## BAN 501

### Module 10 Project

|  |   |
|--|---|
| <b>Project Name</b>  | Supervised Learning – Las Vegas Hotel Reviews |
| <b>Project Due Date</b>  | Sunday by 11:59pm                             |
| <b>Requirements</b>  |   |
| <p>In this project, you will use several techniques of supervised learning to try and predict the number of helpful votes a review receives based on the number of other reviews given by the user.</p> <p>Requirements:</p> <ol style="list-style-type: none"><li>1. Read in the dataset. Note: This csv file is semi colon-separated. You must add the modifier <code>sep=';'</code> as an argument in your <code>read_csv</code> method.</li><li>2. Generate descriptive statistics and print a comment on what you learn about <code>num_reviews</code> and <code>helpful_votes</code> from the statistics.</li><li>3. Save all the <code>num_reviews</code> and <code>helpful_votes</code> to two separate lists.</li><li>4. Using the various sklearn modules, conduct the following analyses to determine the predicted number of <code>helpful_votes</code> based on a user who has posted 300 reviews (<code>num_reviews</code>). Print a prediction of the number of helpful votes for each analysis:<ul style="list-style-type: none"><li>• linear regression</li><li>• nearest neighbor (use <code>n_neighbors=15</code> for the regressor option)</li><li>• decision tree (use <code>max_depth=3</code> for the regressor option)</li><li>• random forest</li><li>• neural network</li></ul></li><li>5. Print a comment about what you learned from conducting the various analyses. Which do you think is best for this situation?</li><li>6. Change the number of reviews from 300 to 10. Run the analyses again and print a comment on what you learned when comparing a small number of reviews to a large number of reviews.</li><li>7. Calculate and print mean prediction error scores for each of the structured learning analyses. (You should still use 10 as the number of reviews for this step.)</li><li>8. Create a 75%/25% train-test split of the data. Train the models then test them using the test data. Calculate and print the error (prediction) scores for each model.</li><li>9. Print a comment about what you learned from comparing the prediction scores. Include a comment on which analysis technique is best for this data and situation.</li></ol> |   |
| <b>Completion</b>  |   |
| <p>Successful completion of this program and all the requirements will result in high marks. You are welcome to add additional functionality and to utilize your creativity in making the program even better.</p>   |   |
| <b>Deliverable</b>   |   |
| <p>Submit your Python file to Canvas.</p>  |   |