Requirement This assignment incorporates several elements of the BEAM079 module and requires that you perform an empirical investigation and discussion of the following scenario - building upon the work you may have carried out in BEAM078 Applied Empirical Accounting and Finance. In 2015/2016 the US market saw a record high number of Chapter 11 Bankruptcy filings across the mining industry – unprecedented in recent times. 1, 2 Using financial data, sentiment analysis, and neural networks, your task is to provide empirical analysis and establish whether these bankruptcies could have been predicted, utilising a model that can be used to predict mining industry bankruptcies one year prior to failure. Incorporating a mixture of financial and textual data, you are tasked with analysing the annual reports of 27 bankrupt mining companies, and 73 non-bankrupt companies (the training sample) during the years 2015/2016. The list of companies, the annual reports, and some financial data has been provided3 . Your primary analyses will be based on this “training” sample of 100 firms. Using your python skills acquired within this module (and elsewhere), your report should seek to provide (but not limited to): An introduction – Discussion of the mining industry and recent bankruptcies, what are the potential causes, and what are your research goals? A literature review – You have 3 strands of literature to discuss namely: Bankruptcy prediction literature, sentiment analysis literature, neural network literature. Remember that a literature review should be a cohesive discussion of extant literature, how it relates to your research, and is the main indicator of whether you understand the subject area or not. Utilise the skills that you have learned within BEAM068. Methodology - What statistical tests/models are you going to perform and why? Data – A description of the data and its origin, including summary statistics of the key variables. Univariate analysis – The inclusion of (but not limited to) tests of difference (t-tests) between the two groups of firms in terms of both financial and textual data (tone and readability); correlation – ensure/demonstrate that there are no extreme correlations that will affect your multivariate neural network; a comparison of bankrupt and non-bankrupt word clouds. Multivariate analysis – Neural Network(s). A predictive model that uses both financial and non-financial data. How accurate is it (ROC AUC)? Some extensions might include - Is it better having textual data in the model rather than just having financial variables? Does it retain its accuracy when applied to the validation sample? How much more accurate is a neural network than a logit model – in terms of both training and validation? Weighted vs unweighted sentiment (advanced). Conclusion – What are your key findings and what are the implications? Python code – A code (\*.py) file which documents each stage of your analysis (uploaded as a separate file). Notes: • The provided annual reports were collected from Perfect Filings (UoE database library). • The provided financial data was collected from WRDS compustat. • The list of failed companies was collected from the Lopucki bankruptcy database.4 • The list of sentiment words from the Loughran McDonald dictionary is provided. Feel free to collect any additional information that you feel is necessary to your study.