1. 大家好，我是CSRD的A+實習生，我叫林暐盛，我的專案主題是CSR文字衍生的策略分析系統以及CSR領域的字詞庫建置。
2. 報告的內容簡單地分成專案的動機、專案執行流程以及未來展望
3. CSR報告書已經成為全球企業的趨勢，從CSR報告書中也能看出企業對於永續發展的各項作為，因此，若能夠分析CSR報告書的分析能力，將能夠幫助我們了解許多資訊，包括瞭解友達有沒有跟上國際的永續趨勢、我們是否比同業做的更多做得更好，以及國際的永續標竿企業有哪些值得我們學習的地方。
4. 然而，以傳統的人工方式分析CSR報告書，有許多的困難處: 最主要的就是以人工分析的方式會參雜人的主觀意識，且每本都由人工觀察十分耗時；最後第三以及第四點，CSR報告書以質性敘述為主，也沒有硬性規定該揭露的細節，如同左方是友達報告書其中一頁的截圖，這樣會使得若要拿數本報告書來做比較，會難以找到比較的標準。
5. 因此，我們會發想，是否能夠用程式模型的量化計算是否能夠幫助我們更有效且精確的分析CSR的報告書。而事實上，外界已經有在進行相關的嘗試，例如左邊的圖，是國內的CSRone永續資訊平台與永訊智庫合作，利用模型分析平台中每月發布的所有文章關注的議題，幫助分析永續議題的變化趨勢；\*而右邊的圖則是美國的永續數據分析公司GlobalAI利用領域關鍵字定義，並定義評分標準，分析某投資組合與標竿企業組合在各項永續相關主題的得分。
6. 回到專案部分，我希望能夠建立CSR領域的文字策略分析系統，而它的主要功能包括:1. 國際趨勢掌握度 2. 高效率的報告書重點項目分析 3. 不同時間的報告書關注議題變化比較4. 報告書關鍵字連結度的主題式查找。
7. 而專案的執行流程大致如下。首先我們會將大量的國際標竿企業報告書中的文字題取出來，並做初步的處理。接著我們會從文字中做CSR領域字詞查找，接著進行相似詞的判斷與定義，把同義但是不同文字的詞歸類在一起。再來就會完成CSR領域字詞庫的建立，也就是第一階段的專案成果。完成後，我們會再將報告書利用字詞庫的對應，做關鍵字的特徵分析，最後再將友達的報告書分析結果與國際標竿企業比較，分析結果。接下來我們將每個步驟做詳細說明與成果。
8. 首先是報告書的文字輸入，本次專案中，我們取用的標竿報告書來自於道瓊永續指數所定義的產業中，與電子製造或是科技業相關的六大產業，其中被評選為銅牌以上企業，例如右下角的三星、HP等，取用的報告書則固定取用2019年發布之英文報告書，來自GRI的資料庫。最終取得了25本標竿永續抱高書來做後續分析的文字來源。
9. 輸入文章之後，接下來兩個步驟是做文字的查找以及定義，首先是有CSR意義的字詞查找，首先從文章中取出出現頻率較高的單詞，這些單詞中，會先透過人工篩選找出具有CSR領域意義的專有名詞直接列入字詞庫，例如””就是有CSR領域意義的單詞。而其餘的字詞會再查找前後詞，如果合併之後具有CSR意義，便會納入字詞庫中。例如””在前後文友找到””與””等，就會被列入字詞庫，在本次專案的結果中，共找到364個CSR領域的相關字詞。第二的部分則是相似詞的定義，需要這步驟的原因主要是因為一義多詞的問題，例如我們今天想知道””有沒有出現在某報告書中，但在報告書中都是用””來表示此議題，若直接搜尋””可能就會認為這本報告書的單位並沒有關注這項議題。因此我們需要查詢並訂義上述CSR相關字詞的相似詞，把指稱相同議題的詞歸在同一類。這部分會使用詞語像量化的模型技術，將質性的詞語化為數值化的向量，並透過向量的數值差距作為字詞意義相似度，來做輔助查找。
10. 在經過上述步驟後，就完成了我們CSR領域字詞庫的建立，字詞庫具體來說的功能是甚麼呢?其實在一開始我們有提到，在做文本的模型分析時，通常都需要將文章轉換成以字詞為單位來表示，也就是分詞的動作，這部分主要是靠程式輔助完成，然而在專業領域上，有些專有名詞是利用多個字詞來表達一個意義的，例如這邊所舉例的句子””，如果是一般的分詞工具，分詞的結果是””，但我們都知道””是一個應該合併表達的議題，所以必須依靠字詞庫告訴程式說這兩個字應該被組合在一起做為一個字詞表示，才能呈現出精確的意義，因此在做CSR文本分析時，此字詞庫是非常重要的存在。而在本次專案，我們完成了129個英文字詞的建置，也依據永續議題中不同的領域進行分類，而每個字詞又有包括其相似詞的集合，是非常完整的資料。此外，除了上述步驟我們提到的英文字詞庫外，我們也用相同的步驟建置了中文的字詞庫，包含了191個字詞。
11. 字詞庫完成後，就能後直接來做報告書的後續分析了，依據我們在前面所提到希望策略系統希望擁有的分析功能，我們一向一向來實現。首先是趨勢掌握度，我將標竿企業與友達自己的報告書利用字詞庫提取出關鍵字與字詞出現頻率，分析關鍵字的出現數量與頻率的差別，左上角的圖，代表在上述三大項永續領域與通用領域中標竿企業提及的議題數量與友達報告書所提及的數量差別，下表則是友達在報告書中未提及到的關鍵議題。而右方雷達圖則是分析標竿企業提到的前十大關鍵議題中，與友達進行詞語頻率的比較。
12. 第二項功能主要是解決報告書分析效率的問題，利用分析系統能夠立刻先分析出報告書中著重的領域以及常提到的關鍵字。以三星的2019年報告書為例，我們就能馬上看出報告書中對於各領域的關注程度，以及三星特別常提及的議題，例如他們很頻繁的提到關於供應商的議題。
13. 第三項功能，則是將不同年份的相同單位報告書輸入，也能夠馬上分析出在不同時期下，關注議題的變化，例如友達在2019年就比2018年更頻繁地提及了能源的相關議題。有了這樣的發現之後，可以後續在透過部門專業的人員去做深入的分析
14. 最後一項功能，主題式的查找，我們可以搜尋一個或是多個關鍵字在文本中出現時，附近通常會提及什麼其他關鍵字或議題，例如說當我們提到TCFD這項議題時，有很高的頻率會伴隨提到風險，那麼他們之間的連結就會越粗而泡泡本身的大小則代表他在整本文本中的出現頻率。
15. 最後，依據本專案的成果，我認為能夠在後續的發展方向，首先是廣泛性，也就是不只利用此系統去分析CSR報告書，而可以將CSR相關的各項文本都作為輸入，例如CSR相關新聞，以增加分析的面向。第二則是實務性，我預計在剩餘的實習期間開發使用者友善的介面，讓後續的研究人員能夠自由且簡單的調整輸入文字以及參數，去進行高效率的運算。這些完成了之後，相性此系統就能夠成為具有高度實用性的CSR策略分析系統，以幫助友達的永續發展訂定準確的方針。