D2K Learning Lab Project

Automated Qualitative and Quantitative Summarization and Visualization of Financial Statements and Corporate Disclosures

Project Pitch

Financial data and company information is often hidden behind layers of fancy terminology, numeric ratios, and the guise that financial advisors are needed in order to make good investment decisions. The SEC requires companies to disclose these values in their 10-Q quarterly reports, but these are often over 50 pages and are not written for the average person. Through automatic qualitative and quantitative summarization of financial statements and corporate disclosures, we would like to provide easy to understand summaries through tagging important and useful sentences in the financial reports.

We hope to begin to extract the most useful information from these reports by building a model where important sentences for determining company performance are extracted from the quarterly report. While each report is structured in the same general way, the value in each of these items can vary depending on the activities of the company in the previous quarter. For example, executives tend to keep the market risks of their company as brief as possible, but sometimes, there is information of value in this section. In addition, we can compare companies within a sector of the economy, basing valuable sentences off of a combination of what is often talked about within a certain market.

Tagging and identifying information from the financial reports is valuable due to what can be done with the data. If we are able to identify the key sentences in a quarterly report and extract them, it would then be possible to algorithmically generate a summary of quarterly reports that is standardized and less biased [Some of this work is detailed in the future work section].

Project Description

This project will focus on the identification of key information in quarterly reports. While executive/management summaries of financial reports do detail some of the performance metrics and outlook of the company, much of the important information, related to fundamentals, earnings, and projections, are hidden behind acronyms, ratios, and terminology that makes it increasingly difficult for a layperson to interpret. This project will utilize quarterly reports in the form of 10-Q statements to identify sentences that explain the current state of the company and indications of the future actions of the company. The 10-Q consists of large amounts of bureaucratic jargon, but the data can be mined for the executive prediction for a company.

The financial data we gather, such as the companies' 10-Q reports, will come from the SEC Filings API. Financial statements can be obtained via cURL or python *urllib* requests. There are certain key sections inside the 10-Q reports on which we will perform further analysis. For example, see the sentences below from Apple's quarterly report that we would like to deem as important.

The Company announced iOS 13, macOS® Catalina, watchOS® 6 and tvOS® 13, updates to its existing operating systems, and introduced iPadOS™, a new operating system designed for iPad, all of which are expected to be available in the fall of 2019. The Company also introduced an updated Mac Pro® and an all-new Pro Display XDR, both of which are expected to be available in the fall of 2019.

To achieve this, we will use a combination of clustering algorithms using NLP APIs to organically find information common to many reports along with text classification to isolate sentences that we find valuable.

Data Structure

The SEC requires that all companies file financial reports in a standardized format, such as the 10-Q format for quarterly reports, and the 10-K format for annual reports. The 10-Q follows a specific itemized format: Financial Statements, Management's Discussion, Risk Assessment, and Controls and Procedures. As an example, we attached the table of contents of Facebook Inc's second-quarter 10-Q from 2018:

Specifically, every company includes a qualitative "Management's Discussion and Analysis of Financial Condition and Results of Operations" section after the quantitative financial statements. In these sections, we hope to find key phrases regarding a company's health and recent areas of growth and loss.

Computing and Graphics

Computing and Graphics net revenue of $1.3 billion for the three months ended September 28, 2019 increased by 36%, compared to net revenue of $938 million for the prior year period, primarily as a result of a 10% increase in unit shipments and a 40% increase in average selling price. The increase in unit shipments was primarily due to higher demand for our Ryzen processors, partially offset by a decrease in demand for our Radeon mobile products. The increase in average selling price was driven by higher demand for our Ryzen processors.

Computing and Graphics net revenue of $3.0 billion for the nine months ended September 28, 2019 decreased by 3%, compared to net revenue of $3.1 billion for the prior year period, primarily as a result of lower graphics channel memory sales due to the decline in blockchain-related demand, partially offset by a 16% increase in average selling price. Unit shipments were flat due to higher demand for our Ryzen processors, offset by lower demand for our Radeon channel products caused primarily by the decline in blockchain-related demand and lower sales of our Radeon mobile products. The increase in average selling price was driven by higher demand for our Ryzen processors and datacenter GPUs, partially offset by lower demand for our Radeon channel products.

In AMD's Q3 2019 report, "primarily as a result of lower graphics channel memory sales due to the decline in blockchain-related demand", is something that should be flagged as it details industry trends and how they are managing and accounting for a decline in sales performance.

**ITEM 2. MANAGEMENT'S DISCUSSION AND ANALYSIS OF FINANCIAL CONDITION AND
RESULTS OF OPERATIONS**

Note About Forward-Looking Statements

This report includes estimates, projections, statements relating to our business plans, objectives, and expected operating results that are "forward-looking statements" within the meaning of the Private Securities Litigation Reform Act of 1995, Section 27A of the Securities Act of 1933, and Section 21E of the Securities Exchange Act of 1934. Forward-looking statements may appear throughout this report, including the following sections: "Management's Discussion and Analysis of Financial Condition and Results of Operations," and "Risk Factors." These forward-looking statements generally are identified by the words "believe," "project," "expect," "anticipate," "estimate," "intend," "strategy," "future," "opportunity," "plan," "may," "should," "will," "would," "will be," "will continue," "will likely result," and similar expressions. Forward-looking statements are based on current expectations and assumptions that are subject to risks and uncertainties that may cause actual results to differ materially. We describe risks and uncertainties that could cause actual results and events to differ materially in "Risk Factors" (Part II, Item 1A of this Form 10-Q), "Quantitative and Qualitative Disclosures about Market Risk" (Part I, Item 3 of this Form 10-Q), and "Management's Discussion and Analysis of Financial Condition and Results of Operations" (Part I, Item 2 of this Form 10-Q). We undertake no obligation to update or revise publicly any forward-looking statements, whether because of new information, future events, or otherwise.

OVERVIEW

The following Management's Discussion and Analysis of Financial Condition and Results of Operations ("MD&A") is intended to help the reader understand the results of operations and financial condition of Microsoft Corporation. MD&A is provided as a supplement to, and should be read in conjunction with, our Annual Report on Form 10-K for the year ended June 30, 2016, and our financial statements and the accompanying Notes to Financial Statements (Part 1, Item 1 of this Form 10-Q).

Microsoft is a technology company whose mission is to empower every person and every organization on the planet to achieve more. Our strategy is to build best-in-class platforms and productivity services for a mobile-first, cloud-first world. We develop, license, and support a wide range of software products, services, and devices that deliver new opportunities, greater convenience, and enhanced value to people's lives.

We generate revenue by licensing and supporting an array of software products, by offering a wide range of services, including cloud-based services to consumers and businesses, by designing, manufacturing, and selling devices that integrate with our cloud-based services, and by delivering relevant online advertising to a global audience. Our most significant expenses are related to compensating employees; designing, manufacturing, marketing, and selling our products and services; datacenter costs in support of our cloud-based services; and income taxes.

Highlights from the first quarter of fiscal year 2017 included:

- Commercial cloud annualized revenue run rate* exceeded $13 billion.
- Office Commercial revenue grew 5%, driven by Office 365 commercial revenue growth of 51%.
- Office 365 consumer subscribers increased to 24 million.
- Microsoft Dynamics ("Dynamics") revenue grew 11%, driven by Dynamics CRM Online revenue growth.
- Server products and cloud services revenue grew 11%, driven by Microsoft Azure ("Azure") revenue growth.
- Azure revenue grew 116%, with Azure compute usage more than doubling year-over-year.
- Search advertising revenue, excluding traffic acquisition costs, grew 9%, driven by increased revenue per search and search volume.

* *Commercial cloud annualized revenue run rate is calculated by multiplying revenue for the last month of the quarter by twelve for Office 365 commercial, Azure, Dynamics online, and other cloud properties.*

In June 2016, we entered into a definitive agreement to acquire LinkedIn Corporation ("LinkedIn") for $196 per share in an all-cash transaction valued at $26.2 billion, inclusive of LinkedIn's net cash (the "Merger Agreement"). We will finance the transaction primarily through the issuance of new indebtedness. The Merger Agreement has been unanimously approved by the Boards of Directors of Microsoft and LinkedIn, and has been approved by LinkedIn's shareholders. We expect the acquisition will close in calendar year 2016, subject to satisfaction of certain regulatory approvals and other customary closing conditions. The acquisition is expected to accelerate the growth of LinkedIn, as well as Office 365 and Dynamics.

Shown above is Microsoft's Management Discussion section from their third quarter 10Q from 2016. In this discussion, the report highlights how Microsoft recently "entered into a definitive agreement to acquire LinkedIn Corporation ("LinkedIn") for $196 per share in an all-cash transaction valued at $26.2 billion". We believe that acquisition announcements like these could potentially be correlated with strong current performance and indicative of continued prosperity in the near future.

Project Objectives

1. Obtain quarterly reports from well-known tech corporations relevant to the project and create our datasets. Cleaning and structuring the data (reports, financial) will also be an important part of this step. Similarly, it is important that we have a database that contains information for these sources for many companies and to have explored this data and drawn some high-level conclusions.
2. Determine what information (e.g. keywords) is important to the summarization of the financial reports and utilize NLP to cluster the data, and to algorithmically extract the important phrases/sentences.
3. Create a platform with data-driven decisions (D^3) that allows us to pick a company and receive a list of sentences and phrases from the quarterly report about the performance/future of the company, to allow for more informed investing.

Background Domain Knowledge

Since we are planning to make the information retrieved in this project accessible and important to those with little financial background, we must be able to determine what information in the reports is useful. We have students who have taken financial accounting so they have just enough knowledge to parse the data. If need be, we can contact professors at the Jones Business School to see if we are missing any vital information. On the coding side, we have the knowledge to gather 10-Q forms from the SEC database. The SEC Filings API returns URLs to web pages that link to 10-Q forms, which can be easily retrieved and converted to plain text. However, we will need to gather knowledge on popular NLP techniques, but that can be done through reading the documentation of the python Natural Language ToolKit, or various posts on StackOverflow or towardsdatascience.

Risks

1. Each document is over 50 pages, which can lead to large amounts of data to parse, especially if we are trying to create a clustering algorithm. However, if this is the case, we can always cluster on a subset of the data and extrapolate.
2. The formatting of the documents is different, which may cause slight issues, but as long as all the sections are titled the same way, we are confident we can use these as markers to track where we are in the report.
3. THE NLP algorithms we incorporate may lack the linguistic and semantic understanding of the text because of the use of jargon in financial text.
4. Documents may repeat the same information for the purpose of an introduction or summary. We need to flag keywords with similar meaning as previous identified keywords as not important.

Future Work

Pending the completion of the objectives detailed in our proposal, we have outlined future work that could be completed to make our project a more usable product.
1. Connect the sentences identified together in a way that it can be presented as a summary
2. Extract the quantitative information relevant to understanding a companies performance and provide visualization for the user
   a. Come up with numerical performance metrics that indicate the positive/negative trajectory of a company
3. Add sentiment analysis of Twitter and news articles of the relevant quarter to enhance the understanding of other thoughts and projections of the company.