

Predict Fraud from Financial Text by Machine Learning

According to the survey by Ernst & Young (2010), the fraudulent activity has increased in the post financial crisis year. Therefore, auditors, regulators, and investors are responsible for detecting fraud by choosing both cost-effective and high-quality tools. Studies have shown that publicly listed companies' textual disclosures contain significant information about economic activities and intangibles related to unethical corporate behaviors. There are some scholars conducting several methods to detect the fraud by the language used in the management discussion and analysis. Thus, this project is going to explore the feasibility and accuracy of detecting fraud via several machine learning algorithms. Machine learning method can be a cost-effective fraud-detection tool for professionals in the financial industry. By comparing that effectiveness of our method to alternative fraud detection approaches across our data, we want to figure out the best predicted method to perform the fraud.

Data Source:

The text data is from the firm's annual 10-K filing which is public online, and we will mainly focus on the discussion part, management discussion and analysis (MD&A).

**UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549**

FORM 10-K

☒ **ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**
For the fiscal year ended September 30, 2017
or
☐ **TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**
For the transition period from _____ to _____
Commission File Number: 001-36743


Apple Inc.
(Exact name of Registrant as specified in its charter)

Example: 10-K filing of Apple Inc.

Item 7. Management's Discussion and Analysis of Financial Condition and Results of Operations

This section and other parts of this Annual Report on Form 10-K ("Form 10-K") contain forward-looking statements, within the meaning of the Private Securities Litigation Reform Act of 1995, that involve risks and uncertainties. Forward-looking statements provide current expectations of future events based on certain assumptions and include any statement that does not directly relate to any historical or current fact. Forward-looking statements can also be identified by words such as "future," "anticipates," "believes," "estimates," "expects," "intends," "plans," "predicts," "will," "would," "could," "can," "may," and similar terms. Forward-looking statements are not guarantees of future performance and the Company's actual results may differ significantly from the results discussed in the forward-looking statements. Factors that might cause such differences include, but are not limited to, those discussed in Part I, Item 1A of this Form 10-K under the heading "Risk Factors," which are incorporated herein by reference. The following discussion should be read in conjunction with the consolidated financial statements and notes thereto included in Part II, Item 8 of this Form 10-K. All information presented herein is based on the Company's fiscal calendar. Unless otherwise stated, references to particular years, quarters, months or periods refer to the Company's fiscal years ended in September and the associated quarters, months and periods of those fiscal years. Each of the terms the "Company" and "Apple" as used herein refers collectively to Apple Inc. and its wholly-owned subsidiaries, unless otherwise stated. The Company assumes no obligation to revise or update any forward-looking statements for any reason, except as required by law.

Overview and Highlights

The Company designs, manufactures and markets mobile communication and media devices and personal computers, and sells a variety of related software, services, accessories, networking solutions and third-party digital content and applications. The Company's products and services include iPhone, iPad, Mac, Apple Watch, Apple TV, a portfolio of consumer and professional software applications, iOS, macOS, watchOS and tvOS operating systems, iCloud, Apple Pay and a variety of accessory, service and support offerings. The Company sells and delivers digital content and applications through the iTunes Store, App Store, Mac App Store, TV App Store, iBooks Store and Apple Music (collectively "Digital Content and Services"). The Company sells its products worldwide through its retail stores, online stores and direct sales force, as well as through third-party cellular network carriers, wholesalers, retailers and value-added resellers. In addition, the Company sells a variety of third-party Apple-compatible products, including application software and various accessories through its retail and online stores. The Company sells to consumers, small and mid-sized businesses and education, enterprise and government customers.

Fiscal Period

The Company's fiscal year is the 52 or 53-week period that ends on the last Saturday of September. The Company's fiscal year 2017 included 53 weeks and ended on September 30, 2017. A 14th week was included in the first quarter of 2017, as is done every five or six years, to realign the Company's fiscal quarters with calendar quarters. The Company's fiscal years 2016 and 2015 ended on September 24, 2016 and September 26, 2015, respectively, and spanned 52 weeks each.

Fiscal 2017 Highlights

Net sales increased 6% or \$13.6 billion during 2017 compared to 2016, primarily driven by growth in Services, iPhone and Mac. The year-over-year increase in net sales reflected growth in each of the geographic operating segments, with the exception of Greater China. The weakness in foreign currencies relative to the U.S. dollar had an unfavorable impact on net sales during 2017 compared to 2016. In May 2017, the Company announced an increase to its capital return program by raising the expected total size of the program from \$250 billion to \$300 billion through March 2019. This included increasing its share repurchase authorization from \$175 billion to \$210 billion and raising its quarterly dividend from \$0.57 to \$0.63 per share beginning in May 2017. During 2017, the Company spent \$33.0 billion to repurchase shares of its common stock and paid dividends and dividend equivalents of \$12.8 billion. Additionally, the Company issued \$24.0 billion of U.S. dollar-denominated term debt, \$2.9 billion of euro-denominated term debt and C\$2.5 billion of Canadian dollar-denominated term debt during 2017.

Fiscal 2016 Highlights

Net sales declined 6% or \$19.1 billion during 2016 compared to 2015, primarily driven by a year-over-year decrease in iPhone net sales and the effect of weakness in most foreign currencies relative to the U.S. dollar, partially offset by an increase in Services. In April 2016, the Company announced an increase to its capital return program by raising the expected total size of the program from \$200 billion to \$250 billion through March 2018. This included increasing its share repurchase authorization from \$140 billion to \$175 billion and raising its quarterly dividend from \$0.52 to \$0.57 per share beginning in May 2016. During 2016, the Company spent \$29.0 billion to repurchase shares of its common stock and paid dividends and dividend equivalents of \$12.2 billion. Additionally, the Company issued \$23.9 billion of U.S. dollar-denominated term debt and A\$1.4 billion of Australian dollar-denominated term debt during 2016.

Example: MD&A of Apple Inc.

Since 1982, the U.S. Securities and Exchange Commission has issued Accounting and Auditing Enforcement Releases (AAERs) during or at the end of an investigation against a company, an auditor, or an officer for alleged accounting and/or auditing misconduct. Based on the AAERs, we can obtain our label: 1 – fraud; 0 – not fraud.

Specifically, our project is a classification problem and we need to build a classifier to tell labels.

Method

1. Data Clean: such as merging the information from AAERs and MD&A; removing punctuations and other basic work for next step
2. Data exploratory Analysis: create word cloud; plot hist gram of the number of sentences in each company's AAERs
3. Word2Vect: convert the text information into numerical information
4. Modeling: Apply the dataset to several machine learning method, such as SVM, Neural Network, Random Forrest etc.
5. Comparison of the machine learning algorithms by decision matrix