

EXTRACTION OF FINANCIAL RATIOS AND ANALYSIS USING WEB SCRAPING AND TEXT ANALYTICS

InvestEazy

PROBLEM

Analysing financial reports of a company can be a very tedious job.

Investors can't depend on financial reports alone, as they don't explain the complete picture of the company's performance.

If investor do not understand what financial terms mean, he/she won't be able to read the reports efficiently.

InvestEazy offers a one stop solution, extracting 15 financial ratios, summarizing chairman's statement, and director's speech.

It also extracts 15 latest news articles and summarizes the sentiment towards the company.

OBJECTIVE

Roadmap

We implemented our project in two methods.

In first method, we worked with PDFs but due to certain challenges, we implemented web scraping.

DATA COLLECTION



WORKING WITH PDFs



WEB SCRAPING



TEXT ANALYTICS



SENTIMENT ANALYSIS



GUI IMPLEMENTATION



WORKING WITH PDFs

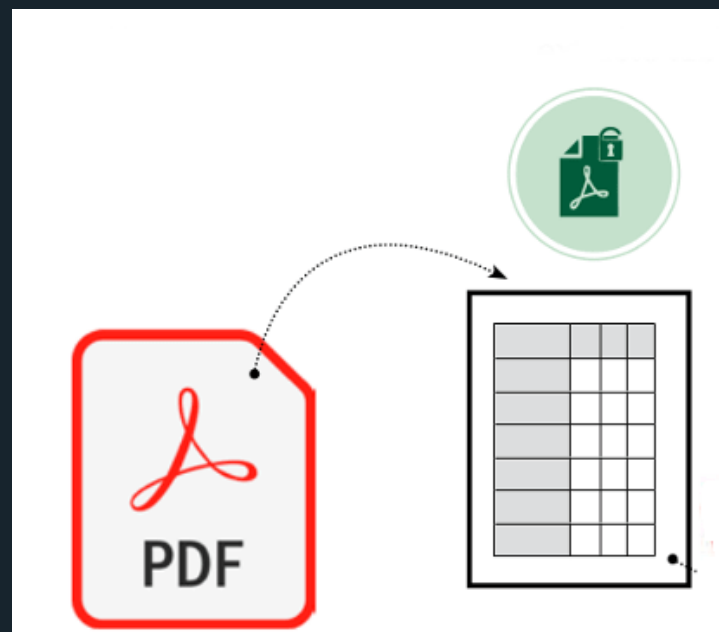
Data Collection:

Collected 3 annual report PDFs each of top 10 Indian Pharma Companies for fiscal years 2017-2019.

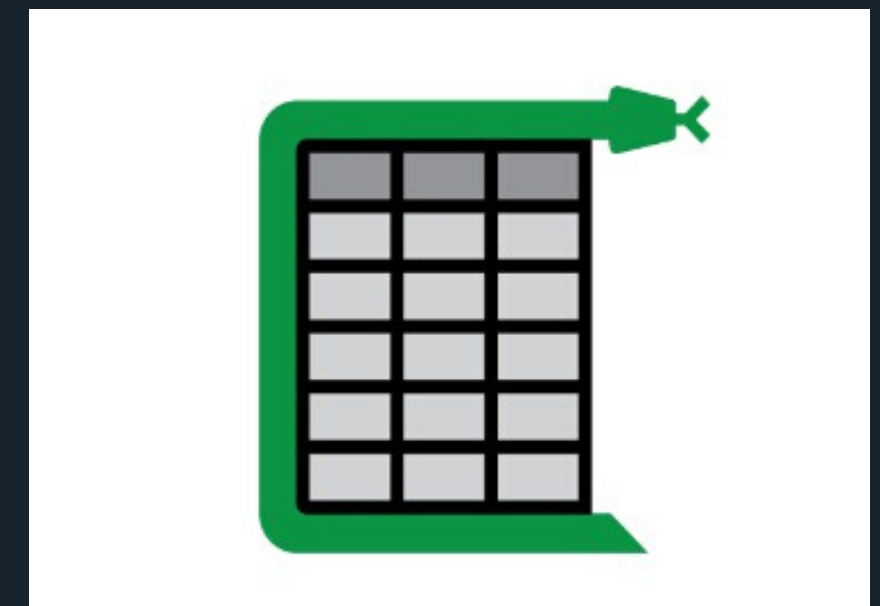
Techniques:



PyPDF2



TABULA



CAMELOT

Challenges in working with PDFs

Unable to extract tabular data in proper format

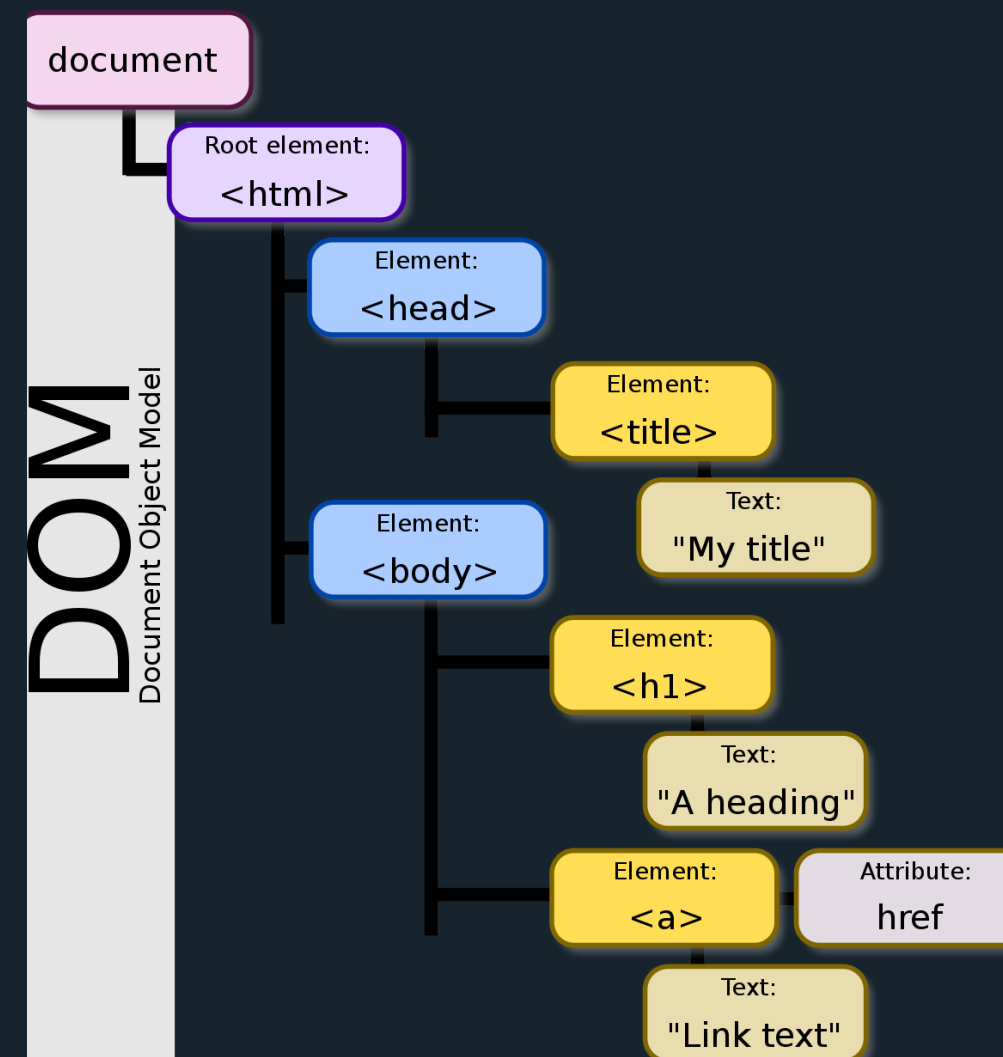
Consolidated and standalone financial statements are detected making multiple instance of statements

Able to work on only annual reports of three years as India implemented new accounting standards

- **Basic Web Scrapping Architecture**



- **Techniques**



Web Scrapping

Extraction of:

- 15 financial ratios
- Chairman's statement
- Director's speech
- 15 latest news articles



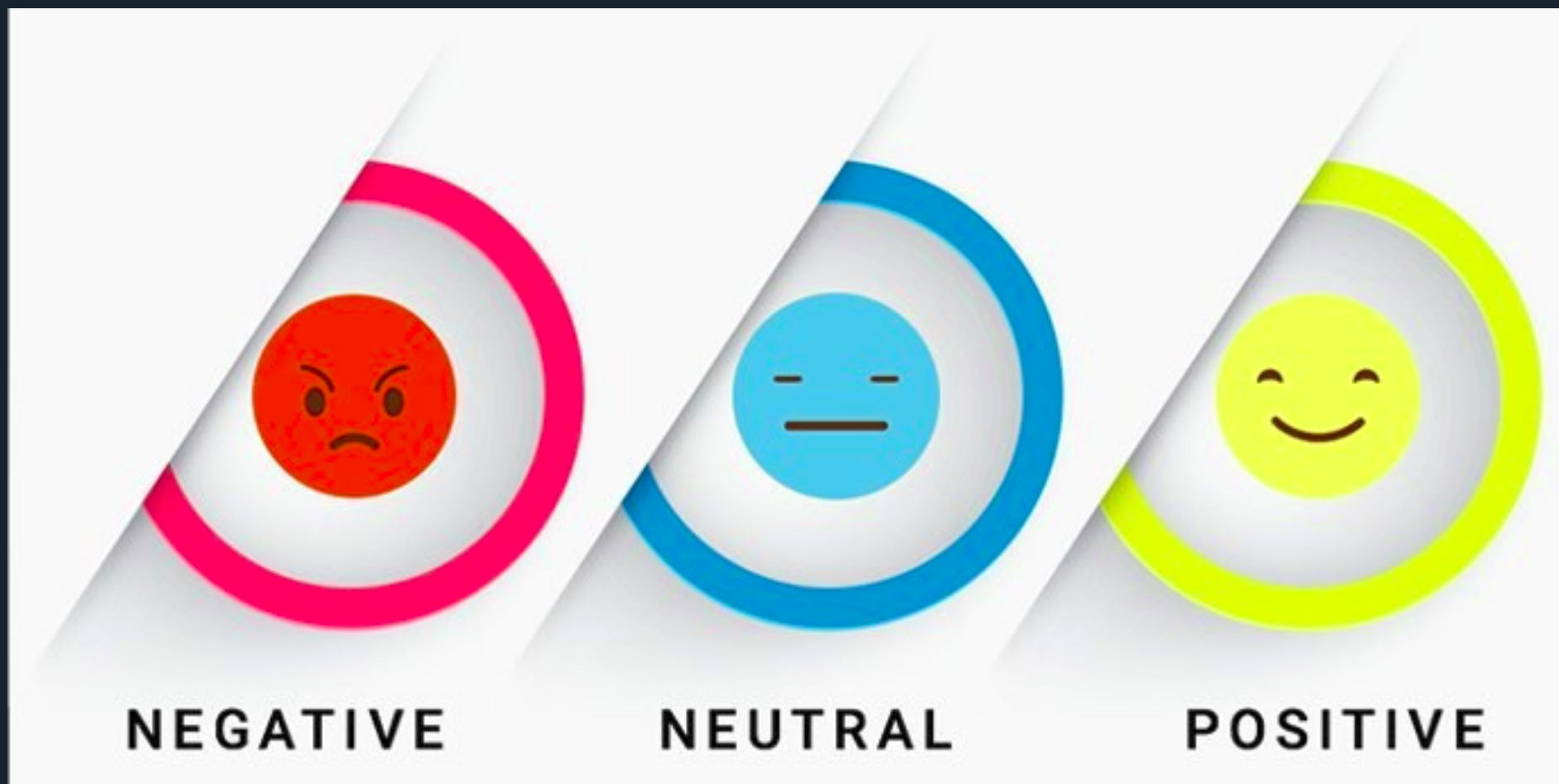
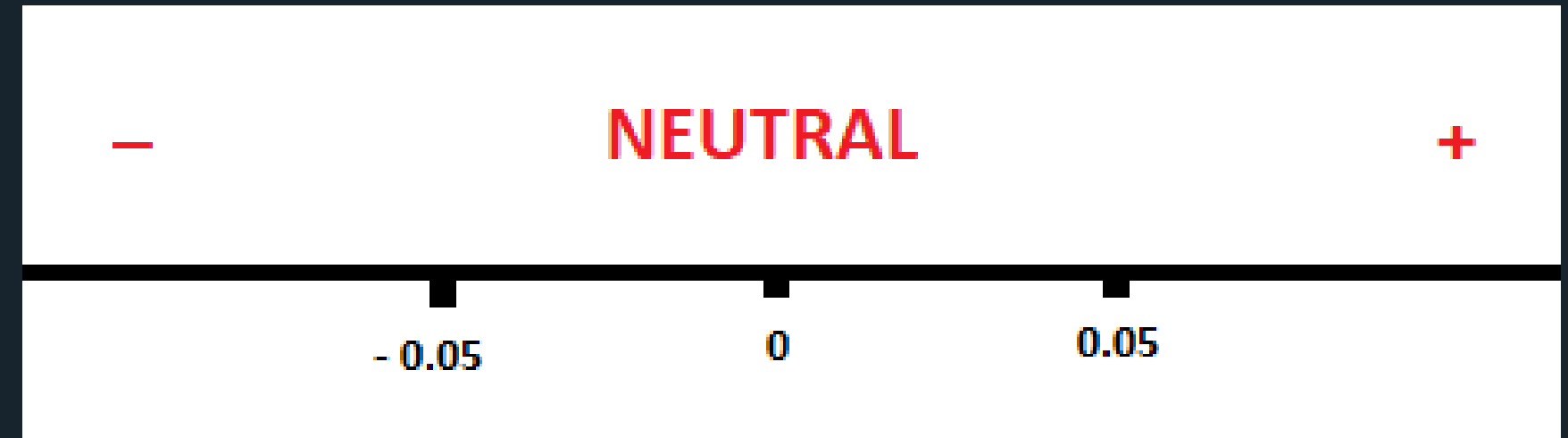
Summarization and Keyword Extraction

- 'Gensim' for extractive text summarization and keyword extraction
- Uses TextRank - graph based ranking algorithm
- Creates summary and extracts keywords from Chairman's and Director's statements



Sentiment Analysis

Uses Vader Sentiment analyzer
for sentiment scores of news
articles



Based on sentiment scores,
general sentiment about
company is derived

GUI: TOOLS

Asian Paints Ltd.											<input type="text" value="Search"/>
BSE 500820											
Ratios	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	See Graph
Basic EPS (Rs.)	80.74	80.81	99.92	109.47	12.19	13.84	16.65	18.78	19.75	22.26	
Diluted EPS (Rs.)	80.74	80.81	99.92	109.47	12.19	13.84	16.65	18.78	19.75	22.26	
Book Value [ExclRevalReserve]/Share (Rs.)	162.35	205.93	259.36	315.08	37.54	44.1	51.74	73.97	81.3	92.66	

CHAIRMAN'S STATEMENTDIRECTOR'S SPEECHNEWS

Chairman's Statement

We continuously innovated by launching new products and services, thereby moving forward in our journey of being the most inspirational home decor brand for consumers.

In such an environment, only those organisations that put in efforts to remain relevant to the changing needs and preferences of consumers - understanding the consumer requirements and adapting the business model to suitably address these requirements, will be able to deliver long-term growth and sustainable returns to all the stakeholders.

As an organisation, we will continue to expand our product offerings and adapt our business channels - from brick & mortar to digital, to deliver on these consumer aspirations.

consumerbusinessproductgenerationchoksi

READ MORE..SHOW LESS..



Future work

- Accommodate financial ratios for banking companies
- Search with company name or BSE number
- Make a database of companies, stock price, director's & chairman's statements, news articles
- Add sentiment for every news article extracted
- Extract news articles from three different websites instead of one website
- Show current and historical data of company's stock price

- By applying web scraping, extractive text summarization technique, and Vader sentiment analysis we were able to develop InvestEazy.
- Unlike moneycontrol, that has data scattered across pages, be it ratios, relevant news, and speeches.
- InvestEazy shows all this information in a single webpage, which could save investors' time and help them make best investments..

CONCLUSION

Our Team



PRAVEEN KUMAR



ABHISHEK KANGALE



VIVEK JOHN P



MOUMITA GHOSH



AKASHNIL ROY



AMEYA KANAWADE

References

<https://github.com/Praveen90179/praxisbusinessschool.capstoneproject6>

