



amazon web services

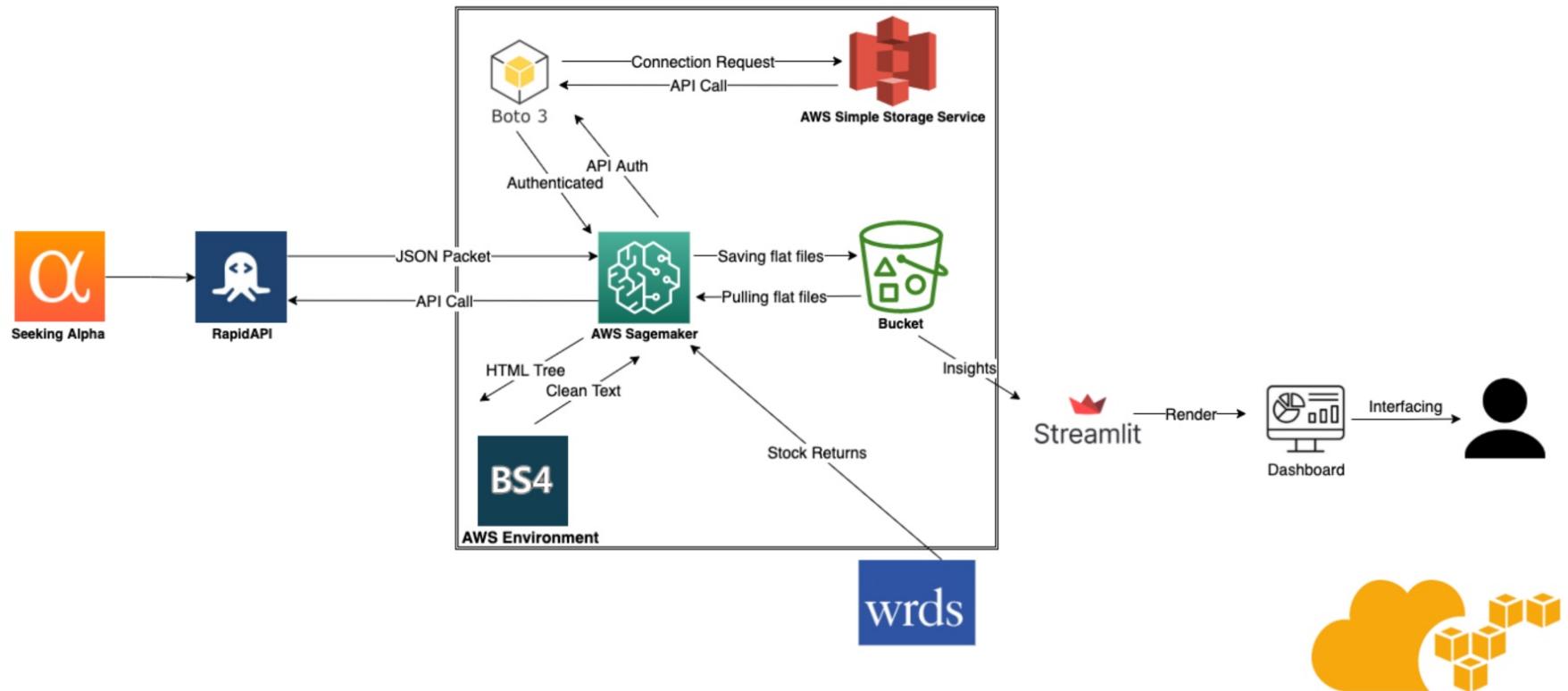


Project Goals/Objectives

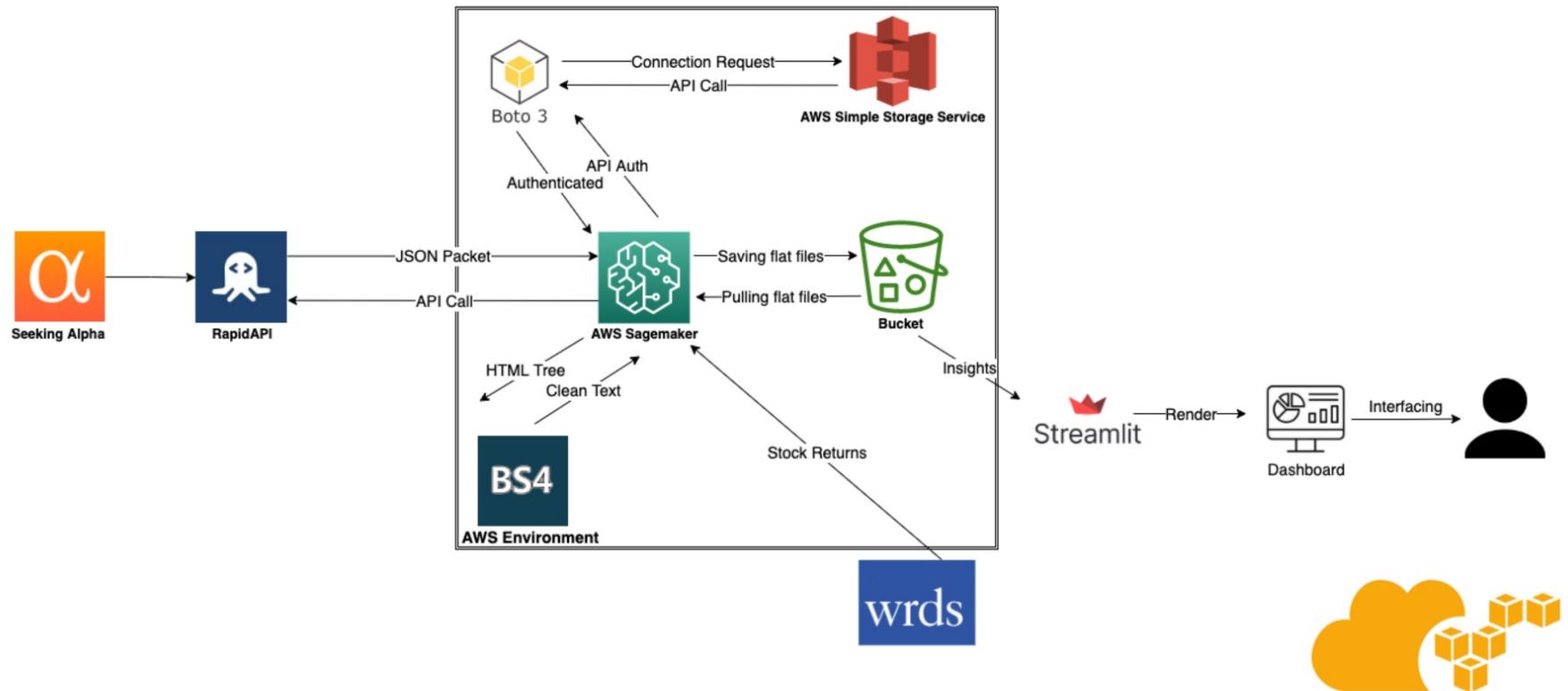
- To analyze the trend in stock prices for tickers of interest over a broad window of time: before the earnings call, during the conference call and after the conference call event
- To obtain earnings call transcripts for the past 10-12 years and analyze them for sentiment scores
- To fit a multimodal machine learning model consisting of earnings call conference texts, corresponding sentiment scores, and stock return data as features; and the stock return data for the month after the actual conference call event as labels.



Architecture



Architecture





Transcripts Data

transcript_name : Name of earnings call

date : Publication date

ticker : Ticker name

clean_text : BS4 Output

company_name : Name of company

id : Unique id of json() packet received

transcript_name	date	ticker	id	clean_text	index	company_name
Agilent Technologies' CEO Discusses F1Q11 Resu...	2011-02-14	A 252805	Agilent Technologies (NYSE:A) F1Q11 Earnings C...	14355	AGILENT TECHNOLOGIES INC	
Agilent Technologies' CEO Discusses Q2 2011 Re...	2011-05-13	A 269881	Agilent Technologies (NYSE:A) Q2 2011 Earnings...	44378	AGILENT TECHNOLOGIES INC	

"Agilent Technologies (NYSE:A) F1Q11 Earnings Call February 14, 2011 4:30 PM ET Executives William Sullivan - Chief Executive Officer, President, Executive Director and Member of Executive Committee Alicia Rodriguez - Michael McMullen - Senior Vice President and President of Chemical Analysis Group Nicolas Roelofs - Senior Vice President and President of Life Sciences Group Ronald Nersesian - Senior Vice President and President of Electronic Measurement Group Didier Hirsch - Chief Financial Officer, Principal Accounting Officer and Senior Vice President Analyst Richard Eastman - Robert W. Baird & Co. Incorporated Nandita Koshal Anthony Luscri - JP Morgan Chase & Co Jonathan Groberg - Macquarie Research Ross Muken - Deutsche Bank AG Stephen Unger - Lazard Capital Markets LLC Isaac Ro - Goldman Sachs Group Inc. Jonathan Palmer - Thomas Weisel Partners D. Mark Douglass - Longbow Research LLC Jon Wood - Jefferies & Company, Inc. William Stein - Crédit Suisse AG Ajit Pai - Stifel, Nicolaus & Co., Inc. Operator Good day, ladies and gentlemen, and welcome to the First Quarter 2011 Agilent Technologies Earnings Conference Call. My name is Derek, and I'll be your operator for today. [Operator Instructions] I would now like to turn the conference over to Ms. Alicia Rodriguez, Vice President of Investor Relations. You may proceed. Alicia Rodriguez Thank you, and welcome, everyone, to Agilent's conference call for fiscal year 2011. With me are Agilent's President and CEO, Bill Sullivan; as well as Senior Vice President and CFO, Didier Hirsch. Joining in our Q&A will be the Presidents of Agilent's Electronic Measurement, Life Sciences and Chemical Analysis Groups: Ron Nersesian, Nick Roelofs and Mike McMullen. After my comments, Bill will give his perspective on the quarter and the overall market result. Didier will then follow with a review of financial results. And after Didier's comments, we will open the lines and take your questions. In case you haven't had a chance to review our press release, you can find it on our website at www.investor.agilent.com. Please note that the business segment financial tables are in the schedules that accompany the press release. We are also providing further information to supplement today's discussion. After you log on to our webcast module from our website, please click on the link for supporting materials. There, you will find additional information, such as our revenue breakouts and historical financial information for Agilent's continuing operations. If during this conference call we use any non-GAAP financial measures, you will find on our website the required reconciliation to the most directly comparable GAAP financial metrics. We will make forward-looking statements about the future financial performance of the company. This involves risk and uncertainties that could cause Agilent's results to differ materially from management's current expectations. As a result, we encourage you to look at the company's most recent filings with the SEC to get a more complete picture of all the factors at work. The forward-looking statements, including our guidance provided today during the call, are only valid as of this date, and the company assumes no obligation to update such statements as we move through

WRDS Stock Return Data



An Affiliate of the University of Chicago Booth School of Business



Daily Stock Return Data- S&P 500 tickers- 2011 to 2021 period

1	PERMNO	date	TICKER	COMNAME	PRC	RET	SHARE VOL
2	10104	1/3/11	ORCL	ORACLE CORP	31.62	0.010224	21136353
3	10104	1/4/11	ORCL	ORACLE CORP	31.48	-0.004428	22978313
4	10104	1/5/11	ORCL	ORACLE CORP	31.04	-0.013977	36464087
5	10104	1/6/11	ORCL	ORACLE CORP	31.17	0.004188	21963429
6	10104	1/7/11	ORCL	ORACLE CORP	31.03	-0.004491	27819266
7	10104	1/10/11	ORCL	ORACLE CORP	31.04	0.000322	29831097
8	10104	1/11/11	ORCL	ORACLE CORP	30.99	-0.001611	26501101
9	10104	1/12/11	ORCL	ORACLE CORP	30.945	-0.001452	34060450
10	10104	1/13/11	ORCL	ORACLE CORP	31.18	0.007594	43574113
11	10104	1/14/11	ORCL	ORACLE CORP	31.245	0.003688	38074998
12	10104	1/18/11	ORCL	ORACLE CORP	31.53	0.009121	28359740
13	10104	1/19/11	ORCL	ORACLE CORP	31.6	0.00222	27094952
14	10104	1/20/11	ORCL	ORACLE CORP	32.305	0.02231	47930038
15	10104	1/21/11	ORCL	ORACLE CORP	32.51	0.006346	34614336
16	10104	1/24/11	ORCL	ORACLE CORP	32.4	-0.003383	20324872

Normalization of stock return w.r.to the market return

- $Mkt - R_f + R_f = Mkt$
- **Normalized stock return** = Stock return - market return
- **ret_d** - stock return for the day
- **ret_normal_d** - normalized stock return
- **ret_mkt_d** - daily market return

[U.S. Research Returns Data \(Downloadable Files\)](#)

[Changes in CRSP Data](#)

Fama/French 3 Factors [TXT](#) [CSV](#) [Details](#)

Fama/French 3 Factors [Weekly] [TXT](#) [CSV](#) [Details](#)

Fama/French 3 Factors [Daily] [TXT](#) [CSV](#) [Details](#)

	F-F_Research_Data_Factors_daily				
	Mkt-RF	SMB	HML	RF	
19260701		0.10	-0.23	-0.28	0.009
19260702		0.45	-0.34	-0.03	0.009
19260706		0.17	0.29	-0.38	0.009
19260707		0.09	-0.59	0.00	0.009
19260708		0.21	-0.38	0.18	0.009
19260709		-0.71	0.44	0.58	0.009
19260710		0.62	-0.52	-0.13	0.009

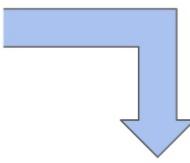
Mkt	RET_percentage	RET_normal
1.18	1.0862	-0.0938
-0.26	-0.9312	-0.6712
0.59	-0.2169	-0.8069
-0.15	0.1932	0.3432
-0.21	0.3375	0.5475

Collection of earnings call dates for tickers

	MMM	AOS	ABT	ABBV	ABMD
0	7/27/2021	4/28/2022	10/16/2014	4/27/2017	4/28/2022
1	7/24/2018	5/05/2020	7/17/2019	2/03/2021	5/04/2017
2	10/27/2024	4/27/2017	1/25/2017	5/01/2020	1/31/2019
3	4/24/2018	4/30/2019	4/16/2020	4/30/2021	5/02/2019
4	10/24/2017	7/29/2021	4/20/2022	4/29/2022	10/26/2019
5	1/28/2020	4/29/2021	1/24/2018	2/07/2020	4/29/2021
6	4/28/2020	1/27/2022	4/17/2019	10/27/2018	8/04/2021
7	1/25/2018	4/25/2018	7/22/2021	2/02/2022	8/01/2019
8	4/26/2022	1/28/2020	1/23/2019	1/27/2017	10/29/2022
9	1/24/2017	10/29/2021	1/22/2020	1/26/2018	2/03/2022
10	4/25/2017	10/29/2011	10/21/2024	4/25/2019	2/06/2020
11	1/25/2022	10/25/2011	1/27/2021	1/25/2019	1/26/2017
12	10/26/2021	1/28/2021	7/20/2017	10/29/2022	7/26/2018
13	10/24/2017	7/30/2019	7/18/2018	7/31/2020	11/01/2011
14	7/25/2017	7/26/2017	4/20/2021	10/30/2022	1/28/2021
15	4/25/2019	7/30/2020	4/18/2018	7/30/2021	10/31/2011
16	4/27/2021	1/29/2019	10/18/2011	11/02/2011	2/01/2018
17	7/28/2020	2/02/2017	10/20/2022	7/28/2017	4/30/2020
18	10/23/2011	10/30/2014	4/19/2017	7/26/2019	8/06/2020
19	1/29/2019	7/25/2018	7/16/2020	4/26/2018	7/27/2017
20	1/26/2021	1/30/2018	10/17/2011	11/01/2015	5/03/2018

date	TICKER	COMNAM	RET	earning_flag
2011/01/03	ORCL	ORACLE CORP	0.010224	0
2011/01/04	ORCL	ORACLE CORP	-0.004428	0
2011/01/05	ORCL	ORACLE CORP	-0.013977	0
2011/01/06	ORCL	ORACLE CORP	0.004188	0
2011/01/07	ORCL	ORACLE CORP	-0.004491	0
...
2016/03/15	CMT	CORE MOLDING TECHNOLOGIES INC	-0.007287	0
2016/03/16	CMT	CORE MOLDING TECHNOLOGIES INC	-0.018760	0
2016/03/17	CMT	CORE MOLDING TECHNOLOGIES INC	0.028263	0
2016/03/18	CMT	CORE MOLDING TECHNOLOGIES INC	-0.001617	0
2016/03/21	CMT	CORE MOLDING TECHNOLOGIES INC	0.000000	0

Marking dates of earnings conference calls as 1



date	TICKER	COMNAM	RET	earning_flag
2012/03/20	ORCL	ORACLE CORP	0.011425	1
2012/06/18	ORCL	ORACLE CORP	-0.020939	1
2012/09/20	ORCL	ORACLE CORP	-0.015863	1
2012/12/18	ORCL	ORACLE CORP	0.017327	1
2013/03/20	ORCL	ORACLE CORP	0.002101	1
...
2020/07/28	ROK	ROCKWELL AUTOMATION INC	-0.029345	1
2020/11/10	ROK	ROCKWELL AUTOMATION INC	-0.022337	1
2021/01/26	ROK	ROCKWELL AUTOMATION INC	-0.062871	1
2021/04/28	ROK	ROCKWELL AUTOMATION INC	-0.008823	1
2021/07/27	ROK	ROCKWELL AUTOMATION INC	-0.014506	1

We obtained a dataframe showing the returns 15 days before and 15 days after every earnings conference call.

	index	date	TICKER	COMMENT	RET	arning_flag	return_d-5	return_d-4	return_d-3	return_d-2	return_d-1	return_d	return_d+1	return_d+2	return_d+3	return_d+4	return_d+5	
0	305	2012/03/20	ORCL	ORACLE CORP	0.011425		1	0.014137	-0.009625	0.007373	-0.010645	0.000673	0.011425	-0.022924	-0.026522	-0.002794	0.021366	0.006516
1	367	2012/06/18	ORCL	ORACLE CORP	-0.020939		1	-0.012887	0.008206	-0.000370	-0.004071	0.029357	-0.020939	0.030973	0.018956	-0.023868	0.006832	-0.013571
2	433	2012/09/20	ORCL	ORACLE CORP	-0.015863		1	0.011159	0.010117	0.004552	-0.002568	-0.007118	-0.015863	0.006510	-0.007546	-0.028704	-0.018371	0.015948
3	493	2012/12/18	ORCL	ORACLE CORP	0.017327		1	0.008419	-0.006803	-0.010332	0.011072	0.011264	0.017327	0.036800	-0.004406	-0.005298	-0.004443	0.000000
4	555	2013/03/20	ORCL	ORACLE CORP	0.002101		1	0.004093	0.020379	0.001102	-0.008255	-0.009711	0.002101	-0.096882	-0.009907	-0.022827	0.009120	0.013160
...	
10026	1046901	2020/07/28	ROK	ROCKWELL AUTOMATION INC	-0.029345		1	0.019736	-0.001014	-0.002692	0.000885	0.012423	-0.029345	-0.005893	-0.005114	-0.007733	0.012056	-0.001450
10027	1046975	2020/11/10	ROK	ROCKWELL AUTOMATION INC	-0.022337		1	0.032520	-0.042152	0.031357	-0.003436	0.021248	-0.022337	-0.027545	-0.029357	0.024545	0.006006	0.000415
10028	1047026	2021/01/26	ROK	ROCKWELL AUTOMATION INC	-0.062871		1	0.021127	0.011379	0.014130	-0.006462	-0.019250	-0.062871	-0.013909	0.042107	-0.010629	-0.003420	0.019541
10029	1047090	2021/04/28	ROK	ROCKWELL AUTOMATION INC	-0.008823		1	0.017476	-0.010343	0.015129	-0.006095	0.004487	-0.008823	0.005709	-0.013108	-0.011693	0.005475	0.014966
10030	1047152	2021/07/27	ROK	ROCKWELL AUTOMATION INC	-0.014506		1	0.019535	0.008009	-0.002615	0.005849	-0.004812	-0.014506	0.019387	0.018818	0.008563	-0.000358	0.012203



Infrastructure as Code

Transcript collection

```
import requests
url = "https://seeking-alpha.p.rapidapi.com/transcripts/v2/list"

headers = {
    "X-RapidAPI-Host": "seeking-alpha.p.rapidapi.com",
    "X-RapidAPI-Key": "ae1546b7b2mshc5ce63ce483644eplc6966jsnb6b90535d501"
}

ids = []
temp_tickers = []
for a in new_tickers.list:
    querystring1 = {"id":a,"until":int(time.time()),"since":"1546304400","size":"40","number":"1"}
    response = requests.request("GET", url, headers=headers, params=querystring1)
    ids.append(response.json())
    temp_tickers.append(a)
    querystring2 = {"id":a,"until":"1546304400","since":"1451610000","size":"40","number":"1"}
    response = requests.request("GET", url, headers=headers, params=querystring2)
    ids.append(response.json())
    temp_tickers.append(a)
    querystring3 = {"id":a,"until":"1451610000","since":"1357002000","size":"40","number":"1"}
    response = requests.request("GET", url, headers=headers, params=querystring3)
    ids.append(response.json())
    temp_tickers.append(a)
    querystring4 = {"id":a,"until":"1357002000","since":"1262307600","size":"40","number":"1"}
    response = requests.request("GET", url, headers=headers, params=querystring4)
    ids.append(response.json())
    temp_tickers.append(a)
```

```
import requests
url = "https://seeking-alpha.p.rapidapi.com/transcripts/get-details"

headers = {
    "X-RapidAPI-Host": "seeking-alpha.p.rapidapi.com",
    "X-RapidAPI-Key": "ae1546b7b2mshc5ce63ce483644eplc6966jsnb6b90535d501"
}

transcripts = {}
clean_transcripts = {}
audio_urls = {}
dates = {}
exceptions = []
for a in temp :
    try:
        querystring = {"id":temp[a]}
        response = requests.request("GET", url, headers=headers, params=querystring)
        transcripts[a] = response.json()
        clean_transcripts[a] = (BeautifulSoup(transcripts[a]['data']['attributes']['content'])).text
        audio_urls[a] = transcripts[a]['data']['transcriptPath']
        dates[a] = transcripts[a]['data']['publishOn'][:10]
    except ValueError:
        exceptions.append(a)
        print(a,temp[a])
    except Exception as e:
        print('New Exception found for :',temp[a],a,e)
```

Infrastructure as Code

AWS S3 Bucket

Saving file from S3 to Sagemaker Environment

```
from sagemaker import get_execution_role
role = get_execution_role()

# Declare bucket name, remote file, and destination
my_bucket = 'sagemaker-studio-owt6bue640l'
orig_file = '20merged_05_16_2022.csv'
dest_file = 'my.csv'

# Connect to S3 bucket and download file
s3 = boto3.resource('s3')
s3.Bucket(my_bucket).download_file(orig_file, dest_file)
```

Saving DF to S3 bucket

```
bucket = 'sagemaker-studio-owt6bue640l' # already created on S3
csv_buffer = StringIO()
final_df.to_csv(csv_buffer)
s3_resource = boto3.resource('s3')
s3_resource.Object(bucket, 'temp_try.csv').put(Body=csv_buffer.getvalue())
```

Code to check files residing in an S3 bucket

```
conn = boto3.client('s3')
contents = conn.list_objects(Bucket='sagemaker-studio-owt6bue640l')['Contents']
for f in contents:
    print(f['Key'])

20merged_05_16_2022.csv
temp_try.csv
```

Infrastructure as Code

To differentiate earnings dates from other dates

```
from collections import defaultdict
earnings_date = defaultdict(set)

for i in transcript.iterrows():
    earnings_date[i[1][3]].add(i[1][1].strip())

returnndf['earnings_flag']=0

for i in returnndf.iterrows():
    date = str(i[1][1]).split()[0]
    ticker = i[1][2]

    if date in earnings_date[ticker]:
        returnndf['earnings_flag'][i[0]] = 1

df_new = returnndf[returnndf.earnings_flag==1]
df_new=(df_new.reset_index())

df_new
```

Infrastructure as Code

Sentiment Scores Calculation

```
print("Starting...")
client = boto3.client('s3')
nlp_score_processor = NLPScorer(
    sagemaker.get_execution_role(),           # Loading job execution role
    2,                                         # instances number, Limit varies with instance type
    'ml.m5.4xlarge',                         # ec2 instance type to run the loading job
    volume_size_in_gb=10,                     # size in GB of the EBS volume to use
    volume_kms_key=None,                      # KMS key for the processing volume
    output_kms_key=None,                      # KMS key ID for processing job outputs
    max_runtime_in_seconds=60*60*18,           # timeout in seconds. Default is 24 hours.
    sagemaker_session=sagemaker.Session(),     # session object
    tags=None)                                # a list of key-value pairs

print("Calculating...")
nlp_score_processor.calculate(
    nlp_scorer_config,
    "clean_text",                            # input column
    's3://earnings-call-transcripts-0/cleaned.csv',   # input from s3 bucket
    's3://earnings-call-transcripts-0/output',          # output s3 prefix (both bucket and folder names are required)
    'scores.csv'                               # output file name
)

print("Downloading...")
client.download_file("earnings-call-transcripts-0", "output/scores.csv", 'output/scores.csv')
print("Done")
```

Infrastructure as Code

Merging of Datasets

```
# First we merge all the returns into a larger dataframe
ret_mkt = pd.read_csv("ret_mkt.csv")
ret_norm = pd.read_csv("ret_norm.csv")
ret = pd.read_csv("ret.csv")

# ret_mkt and ret_norm have the same ordering (sorted by date then ticker), so we merge those first
ret_merged = pd.merge(ret_norm, ret_mkt, left_index=True, right_index=True)

# Now we solve some column name overlaps for all 3 dfs
ret_merged.rename(columns={"return_d": "ret_normal_d"}, inplace=True)
ret.drop(columns=['Unnamed: 0', 'index', 'PERMNO', 'COMMAM', 'PRC', 'RET', 'NUMTRD', 'earnings_flag'], inplace=True)

# ret_merged and ret have different indices, so we sort them both before merging just in case.
ret_merged.sort_values(["TICKER", "date"], inplace=True)
ret.sort_values(["TICKER", "date"], inplace=True)

# Merge mkt/norm with ret
ret_merged = pd.merge(ret_merged, ret, on=["TICKER", "date"])

for i in range(-15, 16):
    if i < 0:
        col = "return_d" + str(i)
    elif i == 0:
        col = "return_d"
    else:
        col = "return_d" + str(i)

    ret_merged = ret_merged[ret_merged[col]!="B"]
    ret_merged = ret_merged[ret_merged[col]!="C"]
    ret_merged[col] = ret_merged[col].apply(lambda x: float(x) * 100)

# Sanity check results
pd.set_option('display.max_columns', None)
mx = 0
for i in range(1000):
    a, b, c = ret_merged.ret_normal_d[i], ret_merged.return_mkt_d[i], ret_merged.return_d[i]
    mx = max(abs(a + b - c), mx)

print("Maximum error in return values:", mx)
if mx > 0.0001:
    raise Exception("Maximum error should be on the order of e-7. Check the scaling for ret_mkt, ret_norm, and ret.")

# Remove duplicate columns
ret_merged.drop(columns=['Unnamed: 0_x', 'Unnamed: 0_y', 'level_0'], inplace=True)
```



Autogluon Modeling & Interpretation

Stock
Return
Analysis

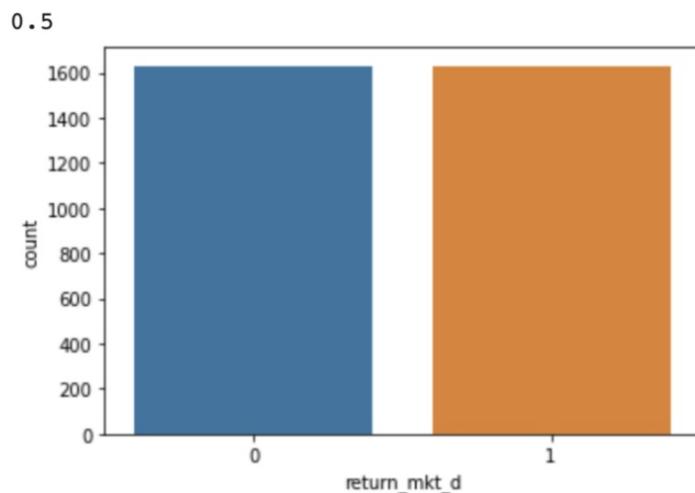
Market
Return
Analysis

Stock Return Analysis

```
df_stock['stock_d'] = df_stock['return_d'].apply(lambda x: 1 if x>=0 else 0)
print(mean(df_stock['stock_d']))
df_stock['stock_0to5'] = df_stock['stock_0to5'].apply(lambda x: 1 if x>=0 else 0)
print(mean(df_stock['stock_0to5']))
df_stock['stock_0to10'] = df_stock['stock_0to10'].apply(lambda x: 1 if x>=0 else 0)
print(mean(df_stock['stock_0to10']))
df_stock['stock_0to15'] = df_stock['stock_0to15'].apply(lambda x: 1 if x>=0 else 0)
print(mean(df_stock['stock_0to15']))
```

```
0.5243357011312813
0.5524861878453039
0.5808997632202052
0.6035253880557748
```

```
import seaborn as sns
sns.countplot(x='return_mkt_d', data=df_market1)
print(mean(df_market1.return_mkt_d))
show()
```



Input
Features

Stock Return Analysis

Input Features

	clean_text	return_mkt_d-15	return_mkt_d-14	return_mkt_d-13	return_mkt_d-12	return_mkt_d-11	return_mkt_d-10	return_mkt_d-9			
0	Agilent Technologies (NYSE:A) F1Q11 Earnings C...	0.65	0.02	0.60	0.26	-1.88	0.71	1.751			
	return_d-15	return_d-14	return_d-13	return_d-12	return_d-11	return_d-10	return_d-9	return_d-8	return_d-7	return_d-6	return_d-5
	1.5673	-0.1871	-0.2811	-0.2819	-3.4629	2.0742	0.5259	-1.9501	-0.5821	4.8792	3.3729
	positive	negative	certainty	uncertainty	risk	safe	litigious	fraud	sentiment	polarity	readability
	0.078651	0.020513	0.028291	0.014486	0.017305	0.041610	0.021291	0.013902	0.103	0.586275	8.83

Models

Stock Return Analysis

Models

Model 1 - Predict the stock return of day 0

```
▶ print('F1 score of training data set:' + str(performance3_2))
print('F1 score of test data set:' + str(perf3_2['f1']))
```

	importance	stddev	p_value	n	p99_high	p99_low
clean_text	0.024798	0.003081	0.002553	3	0.042454	0.007143
return_mkt_d-14	0.007823	0.003795	0.035134	3	0.029566	-0.013921
return_d-3	0.006661	0.003237	0.035253	3	0.025209	-0.011888
return_mkt_d-3	0.006645	0.011302	0.207822	3	0.071409	-0.058118
litigious	0.006223	0.002299	0.021296	3	0.019394	-0.006948
return_mkt_d-15	0.004744	0.003291	0.064949	3	0.023604	-0.014116
return_d-1	0.004638	0.003500	0.074307	3	0.024694	-0.015417
uncertainty	0.004482	0.001971	0.029421	3	0.015775	-0.006812
return_mkt_d-1	0.003814	0.002189	0.047246	3	0.016359	-0.008730

Model 2 - Predict the total stock return from day 0 to day 5

```
print('F1 score of training data set:' + str(performance3_2))
print('F1 score of test data set:' + str(perf3_2['f1']))
```

F1 score of training data set:0.6629732225300092
F1 score of test data set:0.6357039187227866

Autogluon Modeling & Interpretation

Stock
Return
Analysis

Market
Return
Analysis

Autogluon Modeling & Interpretation

Stock
Return
Analysis

Market
Return
Analysis

Market Return Analysis

Model 1: Predict market return of day 0

```
print('F1 score of training data set:' + str(performance2_1))
print('F1 score of test data set:' + str(perf2_1['f1_score']))
```

```
F1 score of training data set:0.964471403812825
F1 score of test data set:0.7862985685071575
```

Model 2: Predict total market return from day 0 to day 5

```
print('F1 score of training data set:' + str(performance2_2))
print('F1 score of test data set:' + str(perf2_2['f1_score']))
```

```
F1 score of training data set:0.9607664233576642
F1 score of test data set:0.7921108742004264
```



Deliverables

- A dataset consisting of text transcripts, stock return data, market return data
- A multimodal machine learning model created using AutoGluon
 - Trained on the dataset mentioned above

[PR FAQ](#)

[Dashboards](#)

PR FAQ

AWS releases a new stock price predicting tool - 'Smart Predictor'

AWS's new stock price predicting tool Smart Predictor analyzes both earnings call transcripts and SEC filings of a company using advanced Machine Learning techniques to make stock price predictions 6 times more precise.

June 2nd, 2022. Today, AWS is excited to announce the addition of a precise and convenient stock price predicting tool Smart Predictor to our customers. The new capability, Smart Predictor, suggests the stock price of a company after an earnings call conference by analyzing both earnings calls transcripts and SEC filings using a neural network model, making it 6 times more precise than the other stock predicting models. For example, if you are interested in the stock price of Apple after its most recent earnings call conference, type its company ticker symbol AAPL as input of the model. Smart Predictor will automatically do the following: scrap the earnings call conference transcripts, extract important information from these files after basic data cleaning and provide a stock price prediction as output. More importantly, Smart Predictor is the first tool analyzing both earnings call transcripts and SEC filings simultaneously for stock price predicting purposes.

Existing stock price prediction solutions analyze earnings call transcripts and SEC filings separately. Analyzing both types of data is a more efficient approach since these files are released quarterly and provide information of a company's value which will be represented in the stock price.

Smart Predictor predicts the stock price with stable and high accuracy in hundreds of tests. It solves the problem by building a multi-layer neural network, which is trained from a huge multi-modal data frame consisting of quarterly earnings call transcripts and SEC filings of hundreds of companies for a 5 years range. It then learns the similarities between the files and stock price and provides a predicted stock price after a new earnings call transcript is input.

"We are excited to bring more value to our customers" said Dr. Bratin Saha, VP of AI/ML Services, AWS, "with Smart Predictor, we are releasing the first of both earnings call transcripts and SEC filings based, deep learning-driven tool to make a more meaningful customer experience."

"Smart Predictor makes it easier for our research work!" said Professor Sanjiv Das, Amazon Scholar, "Smart Predictor has become my team's new favorite tool. After a basic data cleaning process, it extracts the information automatically and gives us reliable predictions!"

FAQs

1. How can I start using Smart Predictor?

You can use Smart Predictor from within SageMaker JumpStart.

2. : What is the price?

This solution is inexpensive to use, you are only billed for the cost of machine usage.

3. Is the result reliable for all the companies?

We trained the model using files of hundreds of companies regarding their earnings call transcripts and SEC filings for a 5 years range. We believe the result is reliable for most of the companies due to the large quantity of our dataset.

Internal FAQs

1. What will customers like the most about Smart Predictor?

Customers will like the increase in efficiency it gives to them, coupled with the general delight in predicting the stock price more precisely.

2. What will customers like the least about Smart Predictor?

Customers will be disappointed in the language coverage along with uncertainties of the model precisionness.
new vers.

3. What are the concerns with this product?

We trained our model based on S&P 500 companies, will the result be reliable for all the other companies? Can we optimize the model to shorten the running time?

4. How do we compare to our competitors?

We are the first to analyze both SEC filings and earnings call transcripts to predict the stock price, the result is more precise since we have more information representing the company's value; We uses a multi layer neural network model trained by a huge data frame, after tuning each of the hyperparameters, the output is of high accuracy in hundreds of test; We create a dashboard which makes it easier for the customers to eyeball the details of input and output.

LITERATURE
REVIEW

Literature Review

Earnings Call Conference, Text Analysis, Seeking Alpha, Orders Backlog, NLP

Earning call conferences are held quarterly by most publicly traded corporations, they usually take place one or two hours before the market opens or closes and sometimes during trading hours, the stock price during these earning calls fluctuate depending on the result and the market expectations. The main reason for earning calls is to instil confidence in investors and analysts and thus the management tries to paint an overly optimistic picture of a company's performance. An earning call lasts for about 1hour, the first 30 minutes are scripted and not much earth shaking news but the analysts Q&A portion can give a lot of information. [Blythe, Bruce](#)

Analysts allowed to participate during earnings conference calls by asking questions receive public signals that can facilitate the generation of new and valuable private information for the asking analyst. Realising these public signals are valuable for the asking analyst, managers can use their discretion to discriminate among analysts by granting more participation to more favourable analysts. Post-Regulation Fair Disclosure conference call transcripts to document that the probability of an analyst asking a question during an earnings conference call is increasing in the favorableness of the analyst's outstanding stock recommendation. [Evidence of Management Discrimination Among Analysts during Earnings Conference Calls](#)

Buy-side analysts are more likely to participate when sell-side analyst coverage is low and dispersion in sell-side earnings forecasts is high, consistent with buy-side analysts participating when a company's information environment is poor.

References

1. Blythe, Bruce. "Quarterly Earnings Calls: Investing Insights from the Source." The Ticker Tape. Accessed December 2, 2021.
<https://tickertape.tdameritrade.com/investing/insights-from-quarterly-earnings-calls-15152>.
2. Amenity Analytics | Use Case | Earnings Call Sentiment Analysis & Text NLP Tool." Accessed December 2, 2021.
<https://www.amenityanalytics.com/case-studies/earnings-call-transcript-analysis>
3. <https://www.sciencedirect.com/science/article/pii/S1090513813000238?via%3Dihub>
4. Yeung, Paper 1 : "Text Mining and Classification on Earnings Call Transcripts." Medium, July 7, 2020.
<https://towardsdatascience.com/text-mining-and-classification-on-earnings-call-transcripts-e63d9856d7d8>.

Deliverables

- A dataset consisting of text transcripts, stock return data, market return data
- A multimodal machine learning model created using AutoGluon
 - Trained on the dataset mentioned above

[PR FAQ](#)

[Dashboards](#)

Dashboard 1

Menu

Company

A

Earnings Call

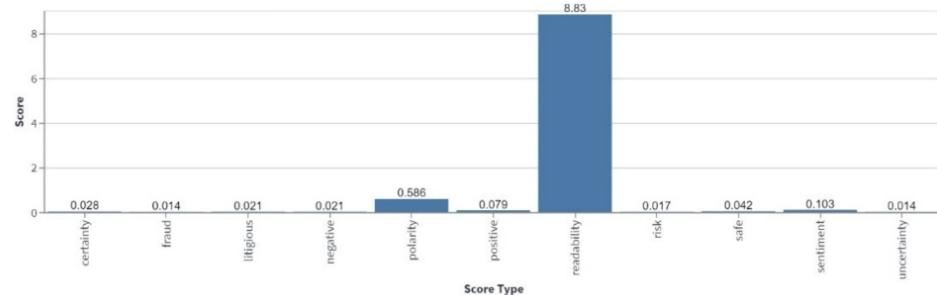
2011-02-14

② A AGILENT TECHNOLOGIES INC

Earnings Call Information

Transcript Name	Agilent Technologies' CEO Discusses F1Q11 Results - Earnings Call Transcript
Call Date	2011-02-14
Stock Price	44.79

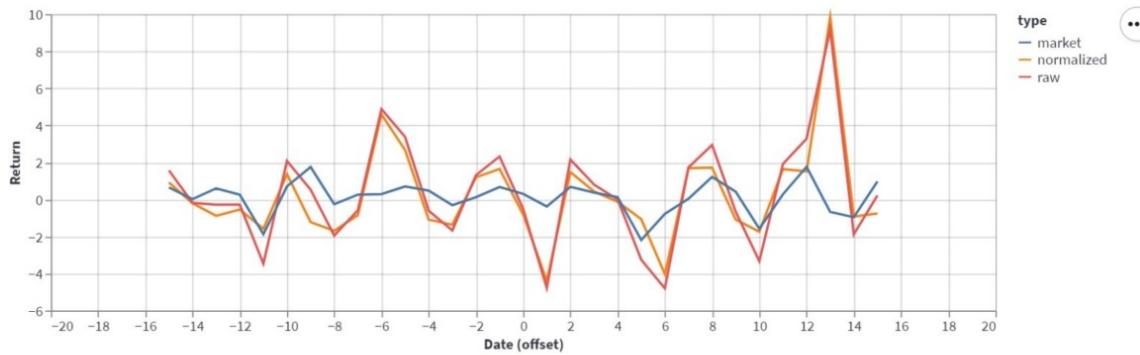
Text Scores



Dashboard
2

Dashboard 2

Returns

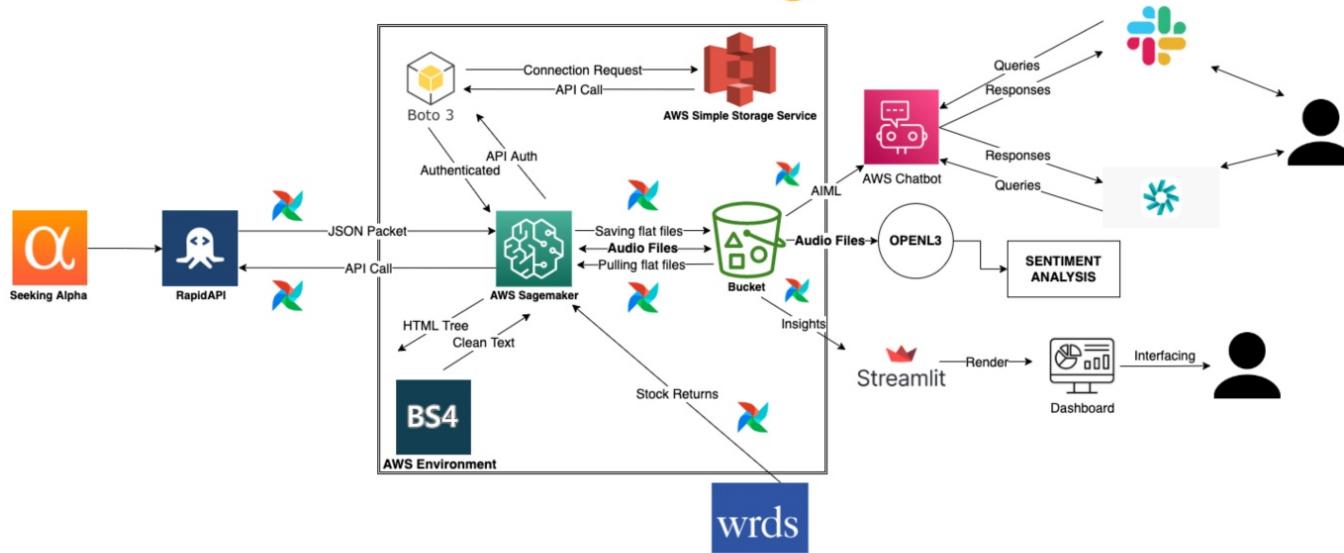


Transcript

Agilent Technologies (NYSE:A) F1Q11 Earnings Call February 14, 2011 4:30 PM ET Executives William Sullivan - Chief Executive Officer, President, Executive Director and Member of Executive Committee Alicia Rodriguez - Michael McMullen - Senior Vice President and President of Chemical Analysis Group Nicolas Roelofs - Senior Vice President and President of Life Sciences Group Ronald Nersesian - Senior Vice President and President of Electronic Measurement Group Didier Hirsch - Chief Financial Officer, Principal Accounting Officer and Senior Vice President Analysts Richard Eastman - Robert W. Baird & Co. Incorporated Nandita Koshal Anthony Luscri - JP Morgan Chase & Co Jonathan Groberg - Macquarie Research Ross Muken - Deutsche Bank AG Stephen Unger - Lazard Capital Markets LLC Isaac Ro - Goldman Sachs Group Inc. Jonathan Palmer - Thomas Weisel Partners D. Mark Douglass - Longbow Research LLC Jon Wood - Jefferies & Company, Inc. William Stein - Crédit Suisse AG Ajit Pai - Stifel, Nicolaus & Co., Inc. Operator Good day, ladies and gentlemen, and welcome to the First Quarter 2011 Agilent Technologies Earnings Conference Call. My name is Derek, and I'll be your operator for today. [Operator Instructions] I would now like to turn the conference over to Ms. Alicia Rodriguez, Vice President of Investor Relations. You may proceed. Alicia Rodriguez Thank you, and welcome, everyone, to Agilent's conference call for fiscal year 2011. With me are Agilent's President and CEO, Bill Sullivan; as well as Senior



Future Scope and Planning



- Continue building on the huge audio database of earnings call conferences.
- Speaker emotion recognition
 - Speaker emotion and sentiment can be classified and added to the data-set
- Make effective use of open source python libraries, such as OpenL3 and Pyannote, for audio file analysis.
- Train the data-set by varying the labels and introducing new features along the way. e.g. use SEC text as a feature.
- Create a continuous-integration and continuous deployment scheme so that the model stays up-to-date with all the new data of the earnings calls and stock returns coming in.
- Building a DAG to automate the pipelines and jobs and triggering them in Apache Airflow
- Developing a question-answering system for the text of SEC filings and earnings calls. Coming up with a list of questions that financial analysts may want to ask about these reports



