STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Dividend Champion Predictor

Zoom Recording Link

AAI 695 - Applied ML Final Project

Connor DePalma, Tim Demetriades, Daniel Pelis

Professor Shucheng Yu

May 6th, 2022

Problem Statement

Dividend Champion

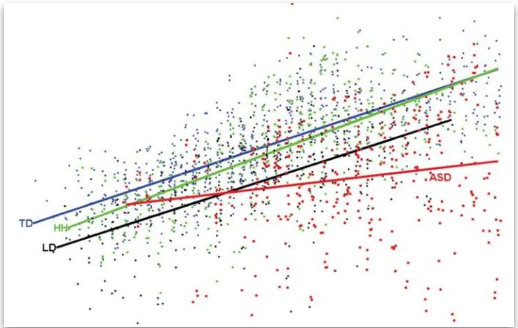
- Companies that have increased their dividends for at least 25 years
- Tend to be some of the most profitable and successful companies
- These companies are will likely perform will even during times of recession
- Some of the safest companies to invest in
- Could be useful to create a system that could predict whether a company will lose its Champion status in the near future

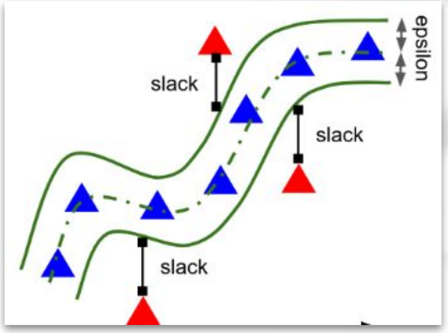
Company	Dividend Yield	Years Dividend Grown
AT&T, Inc.	6.9%	36
Exxon Mobil Corp.	6.1%	38
Chevron Corp.	5.1%	33
International Business Machines Corp.	4.9%	25

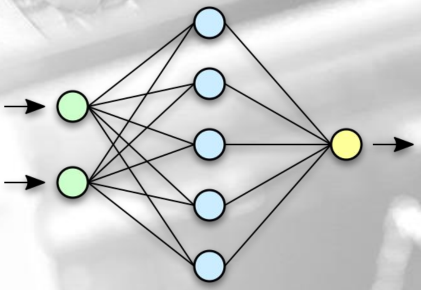
Related Work

Related Work

- Many people have been tracking Dividend Champions in order to understand the best way to make investments
- Many machine learning models have been created that focus on stock price forecasting
  - Multiple regression
  - Support Vector regression
  - Neural Networks
- None of these models tend to focus on the dividends or Dividend Champions which could be very useful





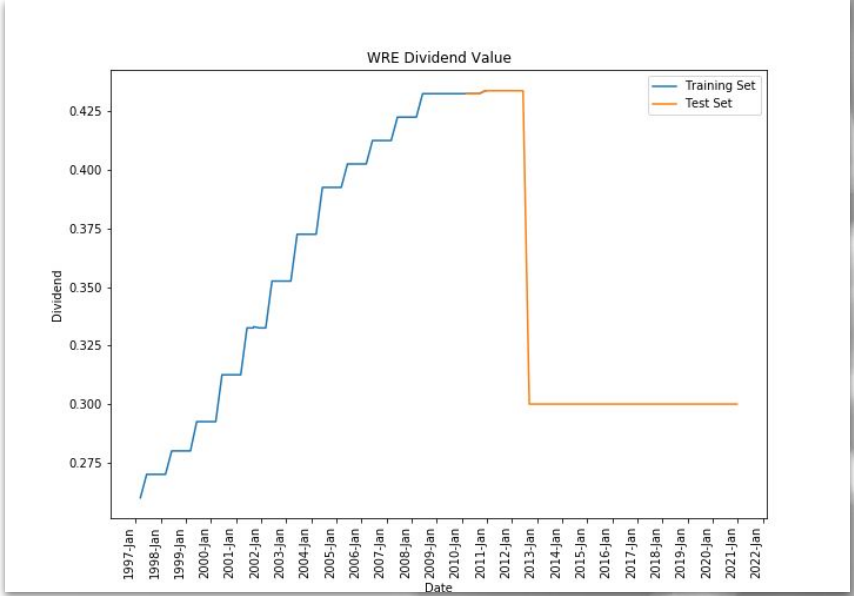


Dataset Description

Dividend Investing Resource Center

- Centralized hub for research and information about dividend investing
- Provides monthly reports of current Champions and their metrics
- Combined with data from the Yahoo Finance API to produce a dataset of past and present Champions

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	*U.S. Dividend Champions*		*End-of-month update at:*							***Dividend Information***			
2	*(and American Depository Receipts)*		http://dripinvesting.org/Tools/Tools.asp							DEG=Dividend to Earnings Growth; A/D=Acceleratic			
3	*12/31/2020*									Dates in Red (right-aligned) indicate last increase mo			
4										Numbers in Gray last updated 12/31/2020			
5	Company	Ticker			**No.**	CCC	DRIP Fees		**12/31/20**	Div.	Current	Payouts/	
6	Name	Symbol	Sector	Industry	**Yrs**	**Seq**	DR	SP	Price	Yield	Dividend	Year	Annualized
7	ABM Industries Inc.	ABM	Industrials	Commercial Ser	**54**	17	Y	N	37.84	2.01	0.1900	4	0.76
8	Archer Daniels Midland	ADM	Consumer Sta	Food Products	**45**	53	N	Y	50.41	2.86	0.3600	4	1.44
9	Automatic Data Proc.	ADP	Information Te	IT Services	**45**	58	N	Y	176.20	2.11	0.9300	4	3.72
10	AFLAC Inc.	AFL	Financials	Insurance	**39**	69	N	N	44.47	2.97	0.3300	4	1.32
11	Albemarle Corp.	ALB	Materials	Chemicals	**26**	126	N	N	147.52	1.04	0.3850	4	1.54
12	A.O. Smith Corp.	AOS	Industrials	Building Product	**27**	123	N	N	54.82	1.90	0.2600	4	1.04
13	Air Products & Chem.	APD	Materials	Chemicals	**38**	70	Y	Y	273.22	1.96	1.3400	4	5.36
14	Arrow Financial Corp.	AROW	Financials	Banks	**28**	107	N	N	29.91	3.48	0.2600	4	1.04
15	Artesian Resources	ARTNA	Utilities	Water Utilities	**28**	108	N	N	37.08	2.77	0.2571	4	1.03
16	Atmos Energy	ATO	Utilities	Gas Utilities	**37**	74	N	N	95.43	2.62	0.6250	4	2.50
17	AptarGroup Inc.	ATR	Materials	Containers & Pa	**27**	120	-	-	136.89	1.05	0.3600	4	1.44
18	American States Water	AWR	Utilities	Water Utilities	**66**	1	N	N	79.51	1.69	0.3350	4	1.34
19	BancFirst Corp. OK	BANF	Financials	Banks	**27**	117	-	-	58.70	2.32	0.3400	4	1.36
20	Becton Dickinson & Co.	BDX	Health Care	Health Care Equ	**49**	39	N	N	250.22	1.33	0.8300	4	3.32
21	Franklin Resources	BEN	Financials	Capital Markets	**41**	64	N	Y	24.99	4.48	0.2800	4	1.12
22	Brown-Forman Class B	BF-B	Consumer Sta	Beverages	**37**	75	Y	Y	79.43	0.90	0.1795	4	0.72
23	Black Hills Corp.	BKH	Utilities	Multi-Utilities	**50**	31	N	N	61.45	3.68	0.5650	4	2.26
24	Badger Meter Inc.	BMI	Information Te	Electronic Equip	**28**	106	N	Y	94.06	0.77	0.1800	4	0.72
25	Brady Corp.	BRC	Industrials	Commercial Ser	**35**	79	N	N	52.82	1.67	0.2200	4	0.88



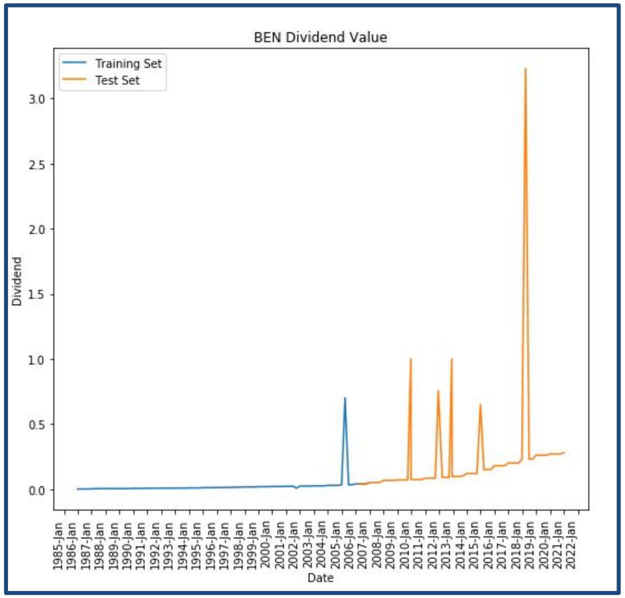
Preparing Training Data

Imperfect Data

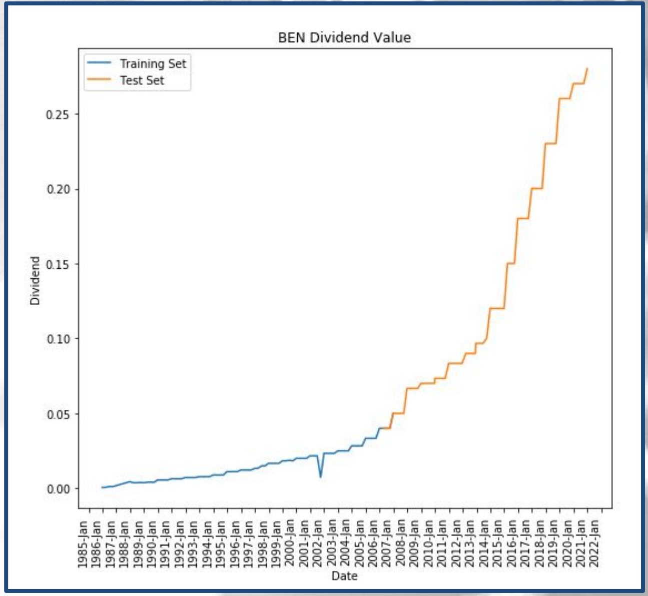
- Not all companies give dividends evenly throughout the year (i.e. every quarter)
- Some companies give "special" dividends that are usually much larger than normal (anomalous samples)

Solution

- Fill in gaps by holding appropriate values so each companies data set is closer in length
- Remove anomalous samples using an edge detection filter and threshold



Remove anomalous samples



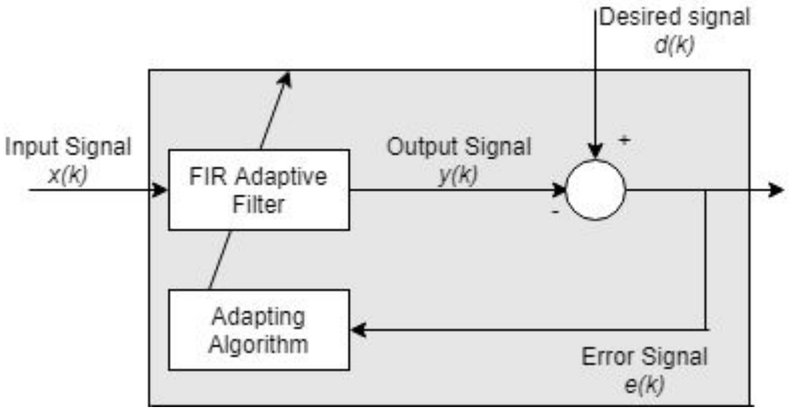
Machine Learning Algorithms

ARIMA

$$\underbrace{y'_t}_{\text{differenced time series}} = \overbrace{c}^{\text{intercept}} + \underbrace{\varphi_1 y'_{t-1} + \ldots + \varphi_p y'_{t-p}}_{\text{lagged values}} + \underbrace{\theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}}_{\text{lagged errors}} + \varepsilon_t$$

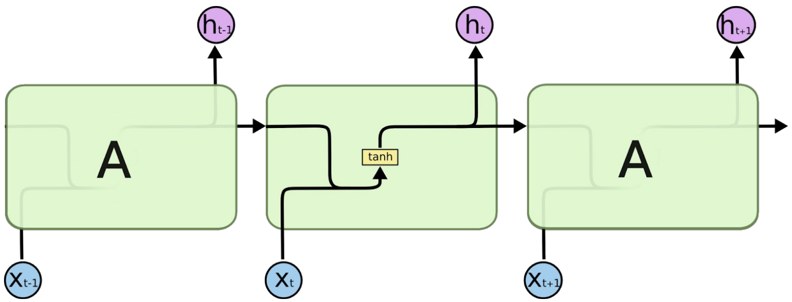
- Simple and popular way to conduct time series analysis/forecasting
- Comprised of 3 parts:
  - **AutoRegressive** (AR) - depends on past values to estimate future values
  - **Moving Average** (MA) - depends on past forecast errors for predictions
  - **Integrated** - data is differenced to make it stationary
- Each part has its own parameter:
  - **p** - order of the AR part, determined by partial autocorrelation (PACF) plot
  - **q** - order of the MA part, determined by autocorrelation (ACF) plot
  - **d** - order of the Integrated part, determined by stationarity test

Adaptive Filters

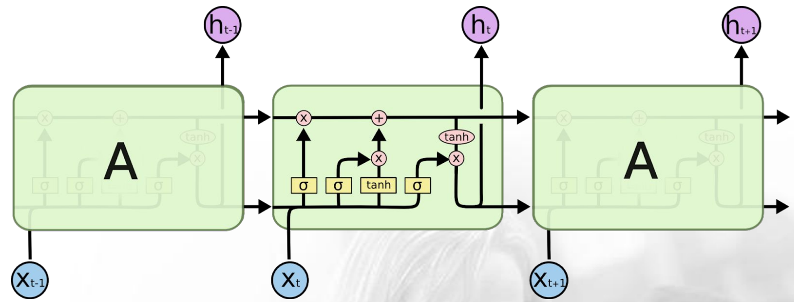


- Digital filters with self-adjusting characteristics commonly used in Image and Signal processing
- Adjusts coefficients to optimize a given cost function
- Least-mean-squares (LMS) is a popular Adaptive Filter that optimizes based on the LMS error of the prediction
  - RLMS is a extension on top of LMS that regularizes the learning rate to provide more accurate and stable results.
- AFs usually take a desired signal represented as a linear equations to use a the ‘actual’ values
  - For Time Series prediction the desired signal is original signal one time step ahead

Recurrent Neural Networks (LSTM)



Classic RNN



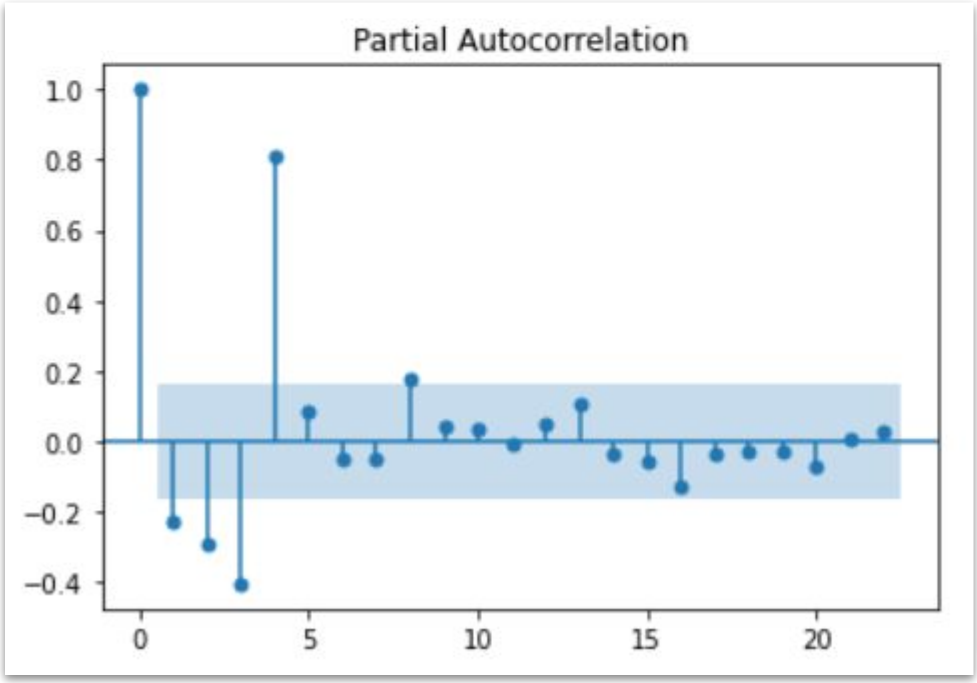
LSTM

- Neural Networks designed to handle sequences of data
- Utilize feedback loops to allow data to be shared across multiple cells (i.e. **memory**)
- Long Short Term Memory (LSTM Networks)
  - Regular RNNs don't have the best memory
    - Only have short term, hidden state
    - Poor long term dependency
    - Often result in Vanishing Gradients
  - LSTM Cells introduce additional long-term memory components called cell state
  - Cell state and hidden state allow for the cell to learn what parts of the sequence to remember and forget, leading to better results compared to classic RNNs

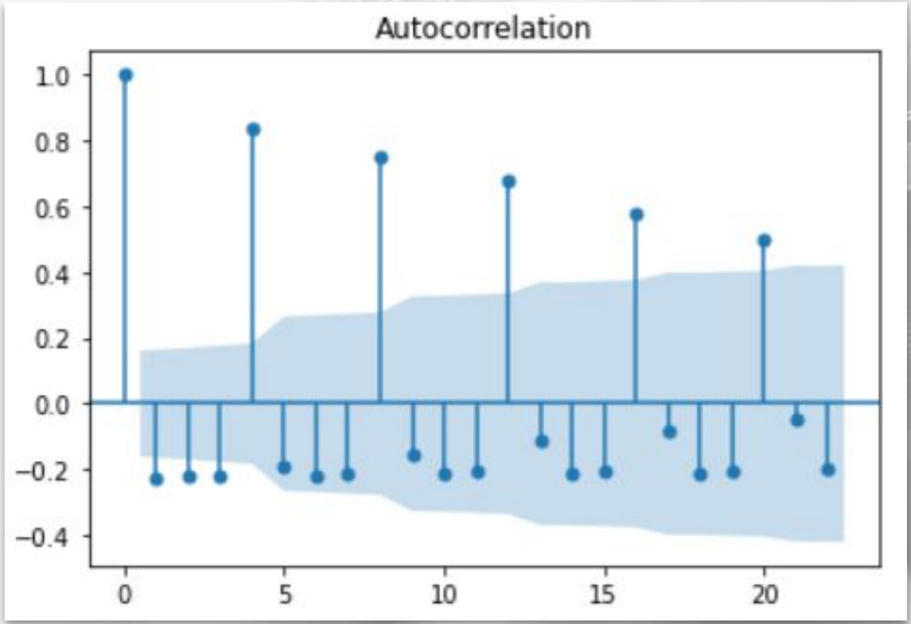
Implementation Details

ARIMA

- To determine order of AR part (p), PACF plot was used:



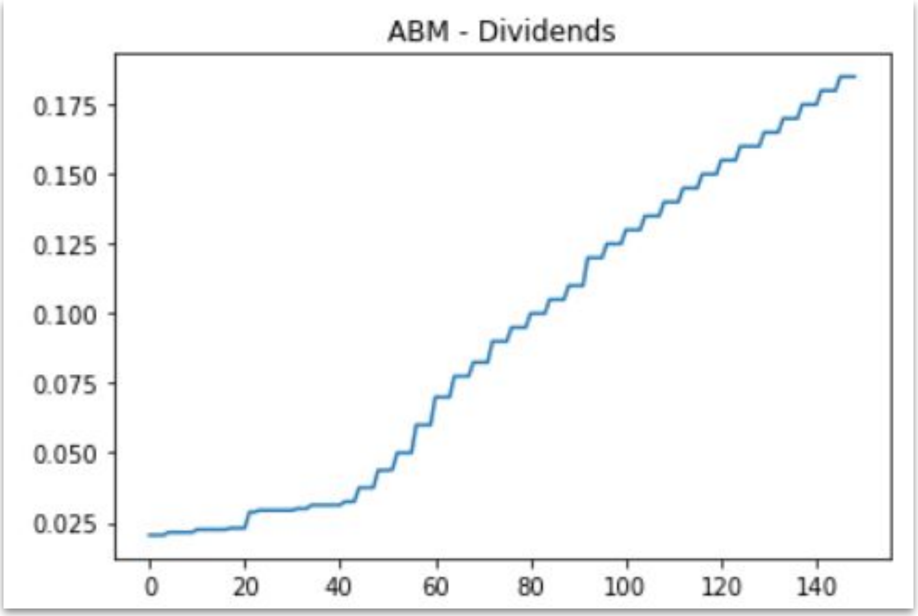
- To determine order of MA part (q), ACF plot was used:

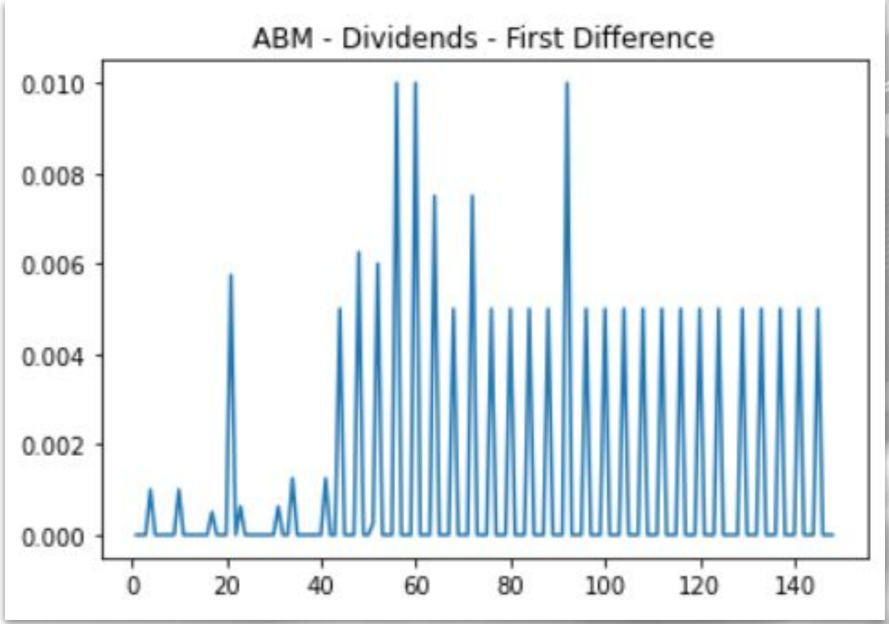


- For the plots above, the optimal order of both the AR part and the MA part of the model would be 4

ARIMA

- As for determining order of the Integrated part (d), this can be gotten by taking d differences of the data and finding how many are needed to make it stationary





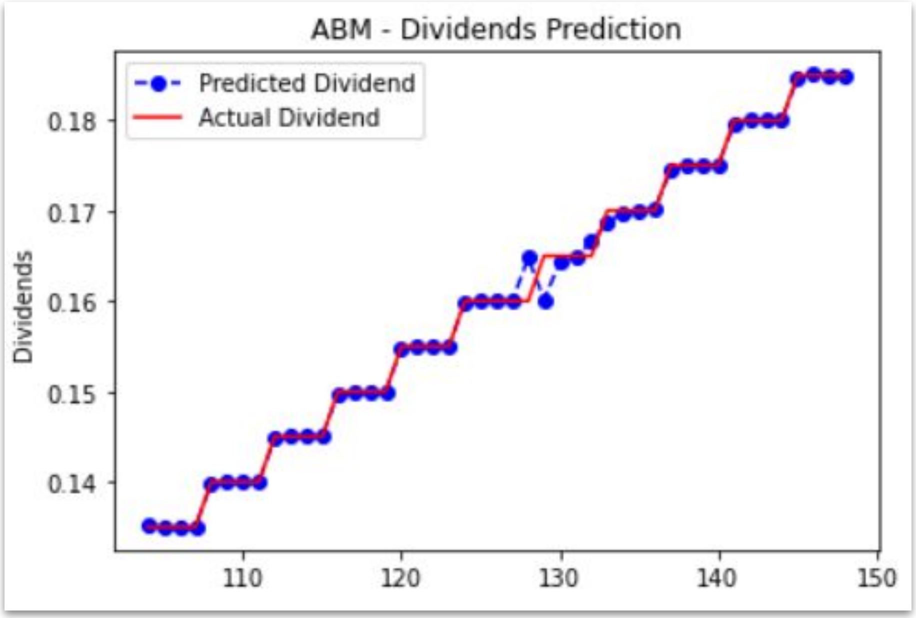

```
ABM
-----
ADF Statistic: 0.649370
p-value: 0.988771
```

```
ABM
-----
ADF Statistic: -3.096103
p-value: 0.026843
```

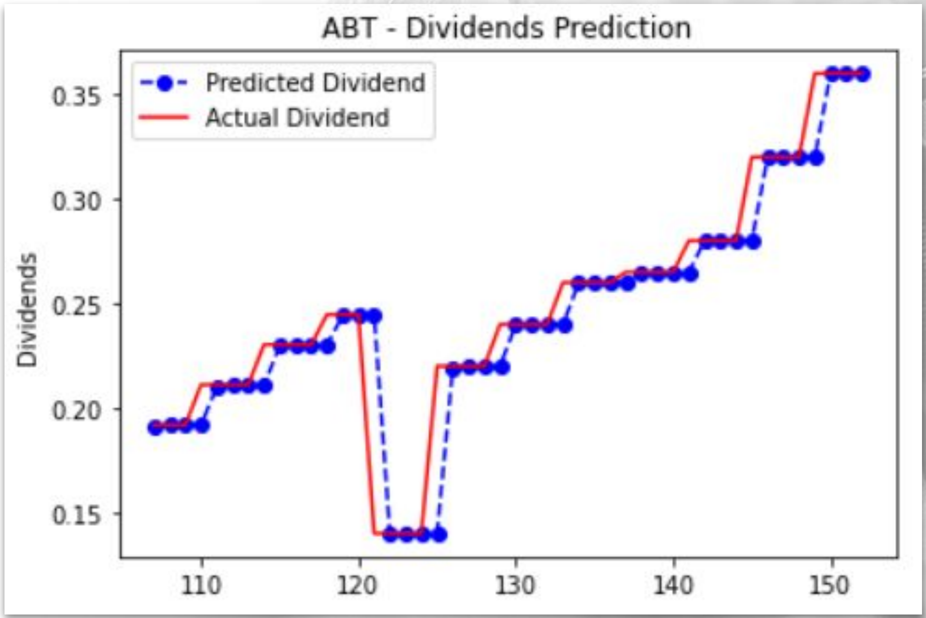
- ADF test can also be done to determine when the data is stationary

ARIMA

- The ARIMA model is able to make accurate predictions when forecasting future values, especially when the order is optimal



ARIMA(4,1,4)



ARIMA(1,1,1)

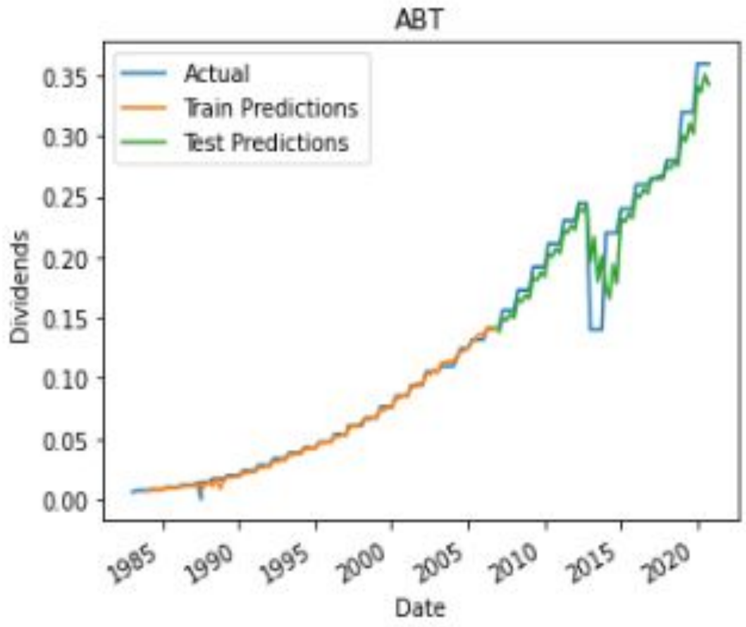
Adaptive Filtering

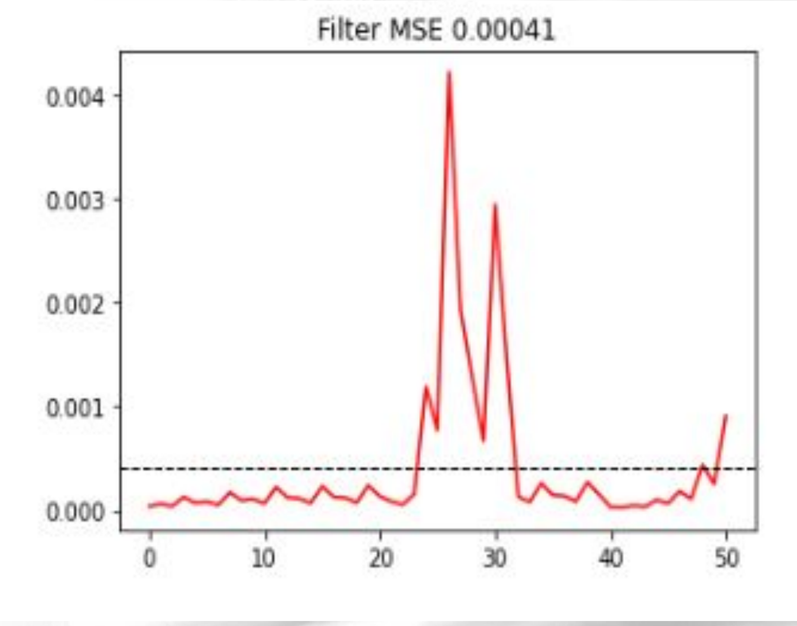
- Padasip
  - Python module designed to simplify adaptive signal processing tasks
  - Functions for input formatting and learning rate tuning
  - Classes for popular adaptive filter types
    - Affine Projection, Least-mean-squares, Recursive Least Squares

- Implementation
  - Number of time steps
    - 5
  - Cost Function
    - NLMS
  - Learning Rate
    - Optimized at runtime

Adaptive Filtering

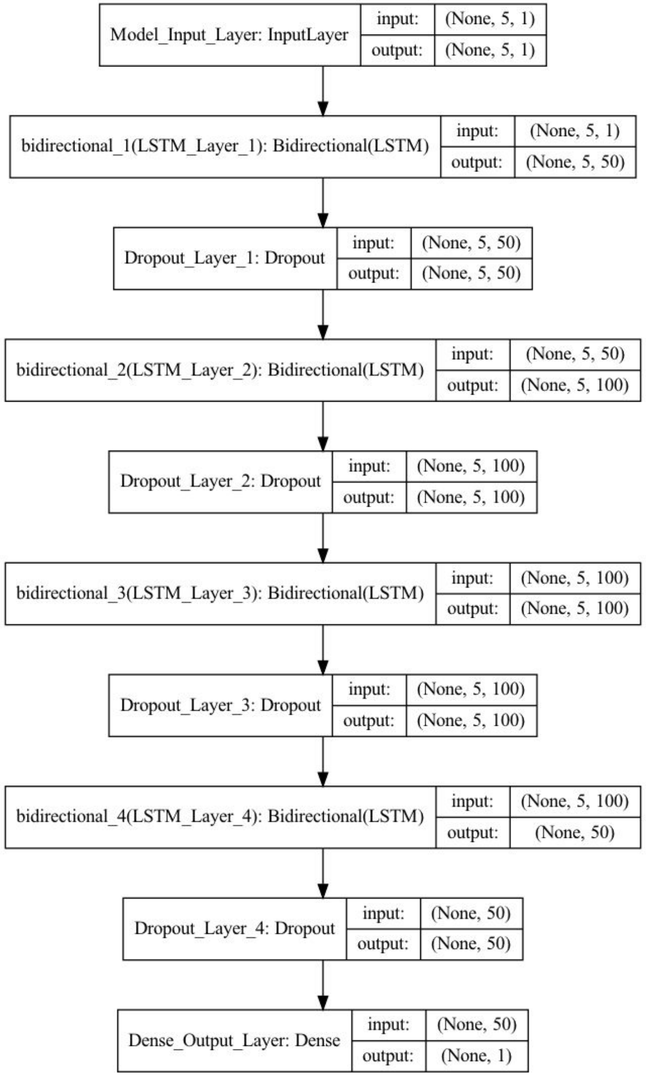
- Tracked input signal well, but does not make drastically jumps
- Random weight initialization leads to varying results





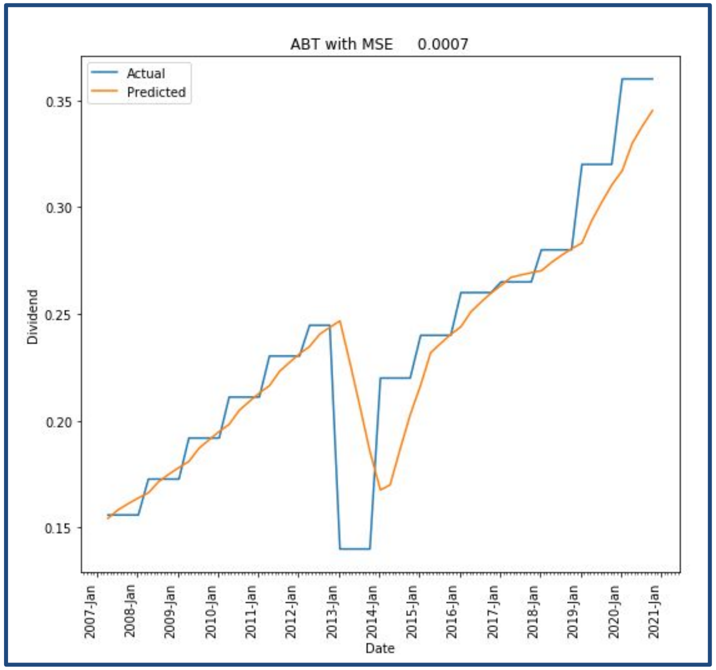
Recurrent Neural Networks (LSTM)

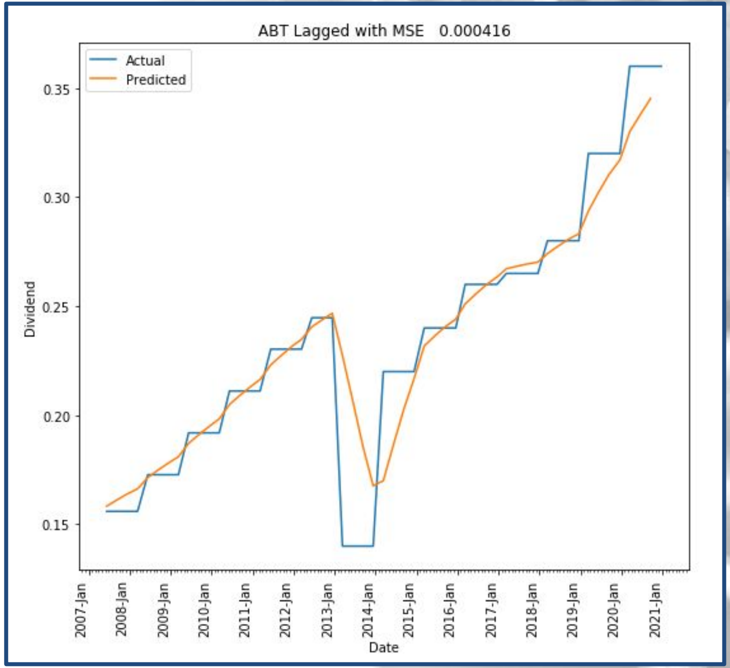
- Model consisted of Input, LSTM, and Dense layers
  - Input layer - Expected a sequence of samples
    - Created by sliding 5 sample long window through datasets
  - LSTM layers - outer LSTM layers had 25 units (output sequence length) and inner had 50
    - Dropout stages were added to statistically mask weights and reduce overfitting throughout the network
    - Bidirectionality was enable to preserve both past and future information and increase accuracy
  - Dense layer - served as output layer with single node
    - This meant the model would take in a sequence of 5 values and output a single prediction for the next value in the sequence



Recurrent Neural Networks (LSTM)

- Hyperparameters were determined through MSE values against validation samples along with visual prediction results
- Visual results were important since both accuracy of the prediction as well as when a significant change in values is predicted need to be considered
  - An incorrect predicted value is not necessarily as important as predicting a the drop in the next dividend value
- Lag adjusted results were also created to show how accuracy could be further improved





Comparison

Comparison

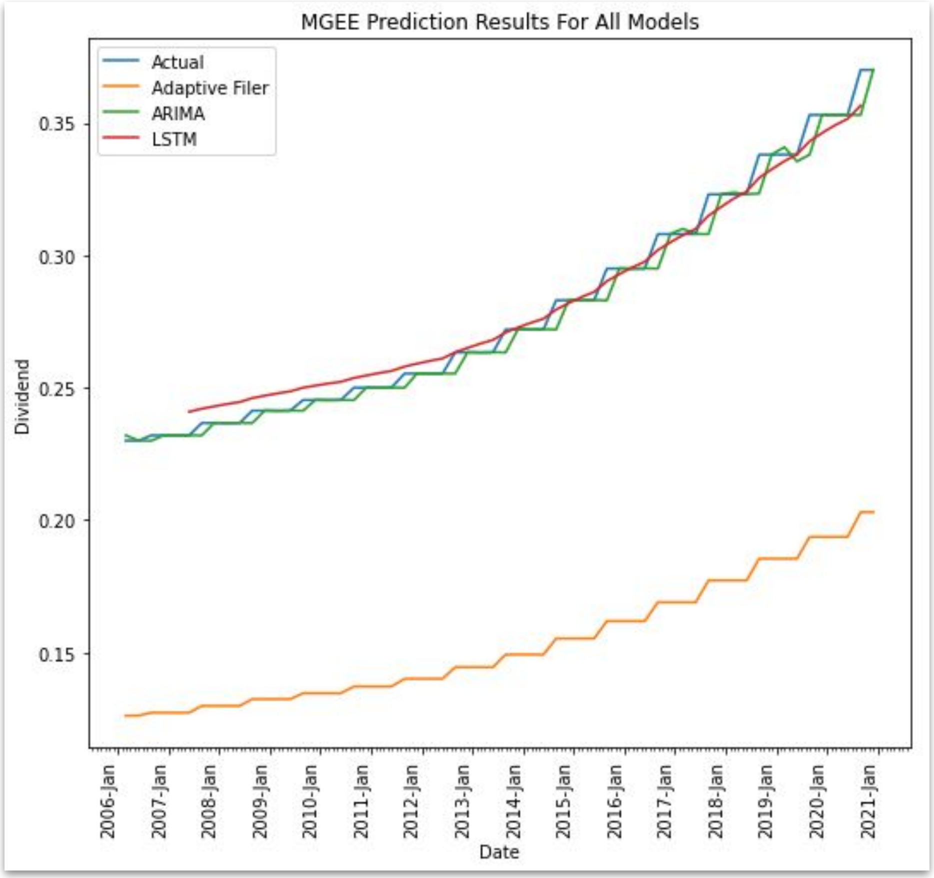
- The below table compares the MSE observed when using each of the three methods for prediction
- 8 companies were chosen - 4 that kept status and 4 that lost status

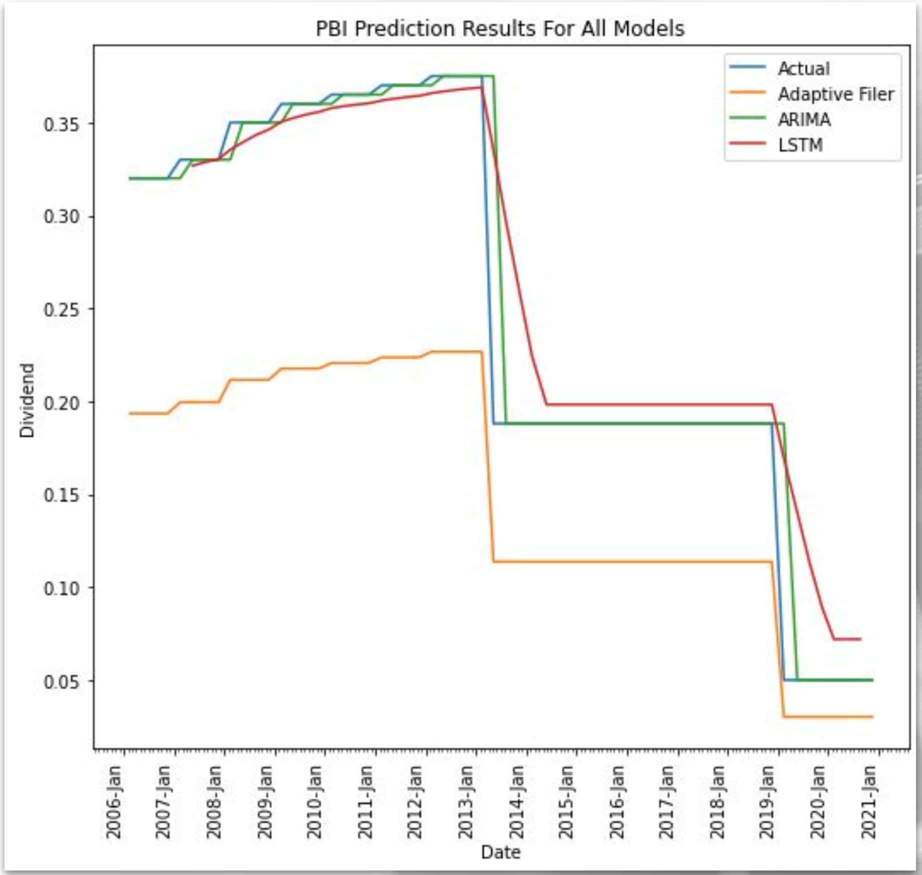
	MGEE	TGT	BEN	CBSH	PBI	ABT	WRE	HP
ARIMA	0.00003	0.00053	0.00009	0.00003	0.00091	0.00039	0.00033	0.00641
AF	0.01609	0.06771	0.01621	0.01126	0.01185	0.01785	0.02700	0.09980
RNN	0.00003	0.00041	0.00009	0.00003	0.00134	0.00042	0.00044	0.00626

- All the values MSE values are relatively low
- ARIMA and RNN have similar and lower values compared to AF
- However AF produced a curve that albeit lower than the actual curve tended to fit its shape better

Comparison

- ARIMA and LSTM better fit to the test set values, but AF better matches the overall shape
- RNN does not have as many predictions as ARIMA and AF





Future Direction

Future Direction

- Models could be improved in various ways
  - More data
  - Better tuning of hyper-parameters
  - Additional testing
  - Ensembling
- Better models could predict more samples, particularly further in the future
- Could also make more accurate predictions
- This would allow for them to be better determiners as for when/if a company will lose Dividend Champion status

Conclusion

Conclusion

- Overall, the Adaptive Filter was decided to perform the best based on its ability to closely follow the shape of test samples
- While the models performed well, they still only predict a single value in the future
    - In order to correctly predict a loss of Champion status, the team would need to make multiple accurate predictions over the course of the year
- Hyperparameter tuning techniques like Gridsearch could be used to further improve the Adaptive Filter model
- Ensembling could be used to combine multiple Adaptive Filter models trained against additional metrics that could be correlated to dividend values

References

- David Enke, Manfred Grauer, Nijat Mehdiyev,
  Stock Market Prediction with Multiple Regression, Fuzzy Type-2 Clustering and Neural Networks,
  Procedia Computer Science,
  Volume 6,
  2011,
  Pages 201-206,
  ISSN 1877-0509,
  https://www.sciencedirect.com/science/article/pii/S1877050911005035

- Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura,
  Stock price prediction using support vector regression on daily and up to the minute prices,
  The Journal of Finance and Data Science,
  Volume 4, Issue 3,
  2018,
  Pages 183-201,
  ISSN 2405-9188,
  https://www.sciencedirect.com/science/article/pii/S2405918818300060

- Chen, D. and Seneviratna, D.
  (2014) Using Feed Forward BPNN for Forecasting All Share Price Index.
  Journal of Data Analysis and Information Processing,
  2, 87-94. doi:
  https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/journal/paperinformation.aspx?paperid=51486