CONTENTS

1. FORECASTING 2
1.1 Forecasting Methods 2
1.1.1 Qualitative Methods 2
a. Buyers intensions survey 3
b. Experts opinion method 3
c. Delphi Method 3
d. Market experimentation 3
1.1.2 Quantitative Methods 3
A. Time Series Models 4
a. Naïve Approach 4
b. Trend Analysis 5
c. Moving Average Model 5
d. Weighted Moving Average Model 6
e. Exponential Smoothing Method 7
f. Holt-Winters Forecasting 7
g. Autoregressive Model 8
h. ARIMA Model 8
B. Associative Models 9
Linear Regression 9
1.2 Defining Forecast Accuracy 12
2. DECISION TREE 15
2.1 Algorithm of decision tree 16
2.2 How to draw a decision tree? 17
2.3 Decision Tree Methods 19
i) CHAID (Chi-squared Automatic Interaction Detector) 19
ii) QUEST (Quick, Unbiased, Efficient Statistical Tree) 19
iii) CART (Classification and Regression Tree) 19
Works Cited 20

1. FORECASTING

Forecasting is the process of making predictions of the future based on past and present data to reduce the risk in all decision making despite many of the decisions are made under uncertainty. Forecasting is beneficial to get a starting point of major decisions in finance, marketing, purchasing and productions. For instance, it is required in predicting demands of new and existing products, projecting quality improvement, predicting cost of materials and anticipating customer`s needs for new product research. In addition to demand forecasts, forecasting is needed for decisions such as identifying labor requirements, anticipating capacity needs, predicting new facility location, developing production schedules…

Why forecasting?

Effective forecasting reduces future uncertainties and builds stability in operations by setting sales targets, pricing policies and establishing controls. Also, it reduces reserves of stock resources to meet uncertain demand by allowing manager to plan personnel, operations of purchasing and finance efficiently. To understand the importance of effective forecasting, forecasting success story of Taco bell appears to be a convenient example that can be given.

Forecasting Customer Demand at Taco Bell: Taco Bell was founded in 1962 and it has grown into an international fast-food business with approximately 6,500 company-owned, licensed, and franchised locations in the United States and around the world with annual sales of approximately $4.6 billion. Labor costs at Taco Bell are approximately 30 percent of every sales dollar, and they are very difficult to manage because labor is closely linked to sales demand. To develop an effective labor management system Taco Bell needed an accurate forecasting model to predict customer demand. Taco Bell determined the best forecasting method was a six-month moving average for each 15-minute interval of every day of the week. The forecasts are compared on a weekly basis to statistical forecasting control limits that are continuously updated, and the length of the moving average is adjusted when the forecasts move out of control. Taco Bell achieved labor savings of more than $40 million from 1993 to 1996 with the new labor management system of which this forecasting model is an integral part.

How?

While forecasting, decision maker should determine purpose of forecast, establish a time horizon and select a forecasting method, respectively. Afterwards, data should be gathered, analysed and finally the forecast will be prepared.

Time Horizon of Forecasting

Long Term: For duration of 3-5 years or more (on annual basis).

Medium Term: For duration of up to 3 years (usually on quarterly or monthly basis).

Short Term: For duration of up to 1 year, usually less than 3 months (on daily, weekly basis).

1.1 Forecasting Methods

1.1.1 Qualitative Methods

Using qualitative approach, a company forecasts based on judgment and opinion**.**

a. Buyers intensions survey

Employs sample survey techniques for gathering data from end users of goods. This method is ideal for short/medium term demand forecasting and cost effective. There is no need for past data so it is the most effective way of assessing demand for new firms and accurate method as buyers` needs are clearly identified. However, people may report what they want to buy, but not what they are capable to buy.

b. Experts opinion method

Panel of exchanging ideas with experts in same area with similar experience and working knowledge. Final decision is based on majority and consensus, reached from expert`s forecasts. Although this method can be undertaken easily without the use of elaborate statistical tools and incorporates a variety of extensive opinions of experts, it has statistical inadequacy since this method lacks statistical data to substantiate the forecast models. Moreover, due to reliance of personal opinions, experts opinion method can be unjustifiable.

c. Delphi Method

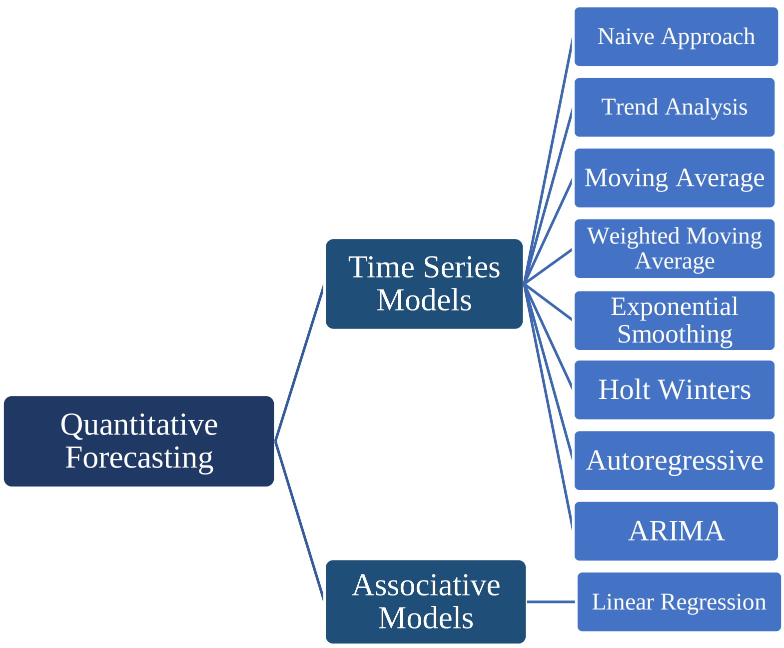
Anonymous forecasts are made by selected experts based on a common questionnaire. Forecasts are revised until a consensus is reached by all. This method is useful for eliminating biases in group meetings but time consuming because reaching a consensus takes a lot of time.

d. Market experimentation

This technique involves actual experiments and proximity with consumers makes collected information reliable. So, it is helpful for estimating actual potential for future sales and improving goods according to provided feedbacks.

1.1.2 Quantitative Methods

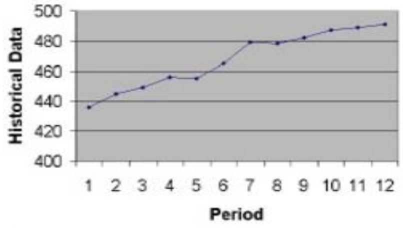
Quantitative forecasting models are used to forecast future data as a function of past data.



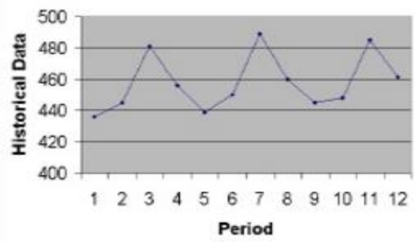
A. Time Series Models

Forecast based only on past values because it assumes that factors influencing past and present will continue influence in future. Time series models have the patterns as following.

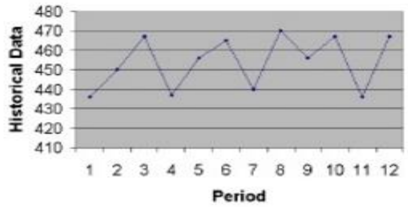
Trend: Demand consistently increases or decreases over time due to population, technology etc. Duration is several years.



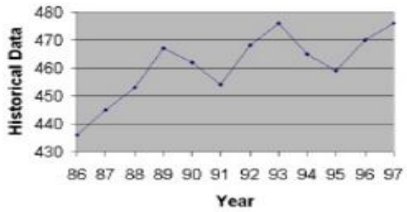
Seasonality: Demand shows peaks at consistent intervals. Regular pattern of up and down fluctuations due to weather, customs etc in 1 year.



Random: Shows unsystematic, residual fluctuations due to random variation or unexpected events such as tornado, union strike. Short duration and nonrepeating.



Cyclical: Demand gradually increases and decreases over time due to interactions of factors influencing economy. Duration is usually 2-10 years.



a. Naïve Approach

Assumes demand in next period will be the same as demand in most recent period. For instance, if June sales were 90, then July sales will be 90 ($y_t = y_{t+1}$, where t represents time). This approach is sometimes cost effective and efficient when the demand is steady or inventory cost is low.

e.g. To observe forecasting methods by performing some applications, the production numbers of Arçelik for washing machine motor in 2017 will be used which are given below as main data.

Motor Krkt.	OCAK	ŞUBAT	MART	NİSAN	MAYIS	HAZİRAN	TEMMUZ
0	69,293	43,503	66,146	41,160	41,765	30,398	22,025
ASKOL	209	0	533	0	0	11	0
ATLAS 22mm	11,857	13,160	21,817	14,517	19,381	13,687	12,892
ATLAS 32mm	82,065	84,991	117,546	124,887	103,622	63,177	80,596
ATLAS 42mm	29,294	21,241	23,819	28,736	29,998	19,704	19,175
ATLAS 50mm	224	393	439	0	511	728	882
DIRECT DRIVE	1,503	1,127	2,096	495	743	1,142	1,801
UNIVERSAL	149,441	144,850	236,981	202,088	158,989	152,312	142,183
Total	343,885	309,265	469,377	411,883	355,010	281,158	279,553

While the production numbers of Arçelik for washing machine motor in 2017, by applying naïve approach to these data it will be as follows.

	Motor Krkt.	OCAK	ŞUBAT	MART	NİSAN	MAYIS	HAZİRAN	TEMMUZ
NAIVE	0		69,293	43,503	66,146	41,160	41,765	30,398
	ASKOL		209	0	533	0	0	11
	ATLAS 22mm		11,857	13,160	21,817	14,517	19,381	13,687
	ATLAS 32mm		82,065	84,991	117,546	124,887	103,622	63,177
	ATLAS 42mm		29,294	21,241	23,819	28,736	29,998	19,704
	ATLAS 50mm		224	393	439	0	511	728
	DIRECT DRIVE		1,503	1,127	2,096	495	743	1,142
	UNIVERSAL		149,441	144,850	236,981	202,088	158,989	152,312
	Total		343,885	309,265	469,377	411,883	355,010	281,158

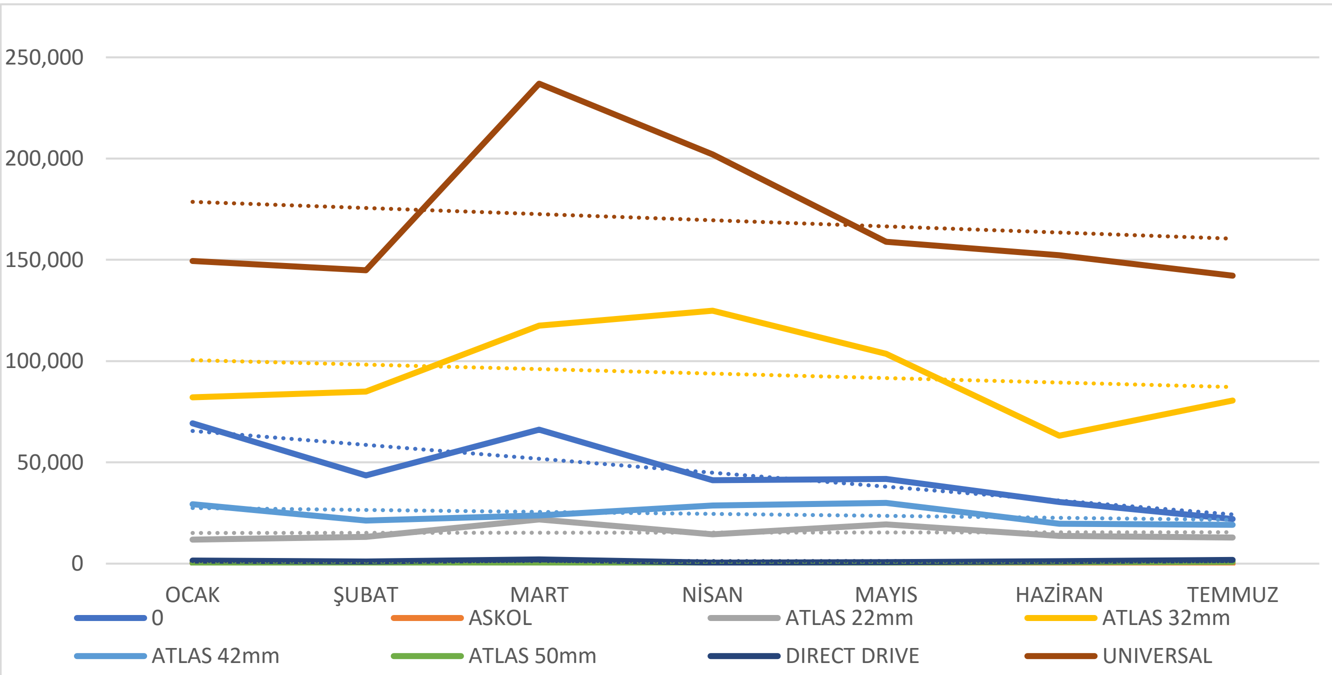
b. Trend Analysis

Product sales, consumer demand and interest rate changes are some of the areas where managers use trend forecasting to determine future profits or losses made by a company. A trend forecast calculates past performance based on a time series that maps specific points within a past trend.

Trend equation: $\boldsymbol{Y^n = a + bx + E}$

(Y^n = Estimated value of Y, a= constant or intercept, b= slope of trend line, x= interdependent variable, E= Error term)

Trend analysis for washing machine motor in 2017 is as following.



c. Moving Average Model

MA is a series of arithmetic means and used if little or no trend, seasonal, and cyclical patterns. Also used often for smoothing since it provides overall impression of data over time.

$$\boldsymbol{MA} = \frac{\sum \boldsymbol{Demand\ in\ previous\ n\ periods}}{\boldsymbol{n}}$$

By applying simple moving average method in production data of washing machine motor in 2017, forecast will be as following.

SIMPLE MOVING AVERAGE

Motor Krkt.	OCAK	ŞUBAT	MART	NİSAN	MAYIS	HAZİRAN	TEMMUZ
0				59,647	50,270	49,690	37,774
ASKOL				247	178	178	4
ATLAS 22mm				15,611	16,498	18,572	15,861
ATLAS 32mm				94,867	109,141	115,351	97,228
ATLAS 42mm				24,785	24,599	27,518	26,146
ATLAS 50mm				352	278	317	413
DIRECT DRIVE				1,575	1,239	1,111	794
UNIVERSAL				177,091	194,639	199,353	171,130
Total				374,176	396,842	412,090	349,350

d. Weighted Moving Average Model

It is used when trend is present (older data is usually less important) and weights are based on intuition (often lay between 0-1).

$$\boldsymbol{WMA} = \frac{\sum(\boldsymbol{Demand\ in\ period\ n}) * (\boldsymbol{Weight\ for\ period\ n})}{\sum \boldsymbol{Weights}}$$

By applying weighted moving average method in production data of washing machine motor in 2017, forecast will be as following.

WEIGHTED MOVING AVERAGE

Motor Krkt.	OCAK	ŞUBAT	MART	NİSAN	MAYIS	HAZİRAN	TEMMUZ
0				59,123	49,879	45,627	35,981
ASKOL				301	178	89	5
ATLAS 22mm				17,271	16,724	18,166	15,723
ATLAS 32mm				100,781	115,791	113,031	86,943
ATLAS 42mm				23,872	25,848	28,548	24,641
ATLAS 50mm				388	212	329	535
DIRECT DRIVE				1,674	1,134	886	901
UNIVERSAL				191,681	204,179	186,354	162,833
Total				395,091	413,945	393,029	327,563

– *Disadvantages of Moving Average Methods:* Increasing n makes forecast less sensitive to changes and the method does not forecast trend well due to the delay between actual outcome and forecast.

Also, it requires much historical data, difficult to trace seasonal and cyclical patterns. So, weighted MA may perform better.

e. Exponential Smoothing Method

A sophisticated weighted moving average method that calculates the average of a time series by giving recent demands more weight than earlier demands. Exponential smoothing requires smoothing constant (α) which ranges from 0 to1. In addition, it is the most frequently used formal forecasting method because of its simplicity and the small amount of data needed to support it.

$\boldsymbol{F_{t+1}} = \alpha(\boldsymbol{D_t} - \boldsymbol{F_t}) + \boldsymbol{F_t}$ or $\boldsymbol{F_{t+1}} = \alpha\boldsymbol{D_t} + (\boldsymbol{1} - \alpha)\boldsymbol{F_t}$ is being used for computing forecast.

(D_t = Demand in period t, F_t = forecasted value in period t)

By applying exponential smoothing method in production data of washing machine motor in 2017, forecast will be as following. ($\alpha = 0.9$)

EXPONENTIAL SMOOTHING

Motor Krkt.	OCAK	ŞUBAT	MART	NİSAN	MAYIS	HAZİRAN	TEMMUZ
0		69,293	46082	64139	43458	41934	31552
ASKOL		209	21	482	48	5	10
ATLAS 22mm		11,857	13030	20938	15159	18959	14214
ATLAS 32mm		82,065	84698	114261	123824	105642	67423
ATLAS 42mm		29,294	22046	23642	28227	29821	20715
ATLAS 50mm		224	376	433	43	465	702
DIRECT DRIVE		1,503	1165	2003	646	734	1101
UNIVERSAL		149,441	145309	227814	204660	163556	153436
Total		343,885	312727	453712	416066	361115	289154

Choosing α: Seek to minimize the Mean Absolute Deviation (MAD)

If: Forecast Error = demand – forecast

Then: $MAD = \sum \frac{|Forecast\ errors|}{n}$

f. Holt-Winters Forecasting

Holt's method can be extended to deal with time series which contain both trend and seasonal variations. i.e. repetitive over some period. The basic equations for the method are given by:

Overall Smoothing: $S_t = \alpha * \frac{y_t}{I_{t-L}} + (1-\alpha) * (S_{t-1} + b_{t-1})$

Trend Smoothing: $b_t = \gamma * (S_t - S_{t-1}) + (1-\gamma) * b_{t-1}$

Seasonal Smoothing: $I_t = \beta * \frac{y_t}{S_t} + (1-\beta) * I_{t-L}$

$$F_{t+m} = (S_t + m * b_t) * I_{t-L+m}$$

Where y is the actual value, $S$is the smoothed actual value, b is the trend factor, I is the seasonal index, F is the forecast at m periods ahead, t is an index denoting a time period and α, β, γ are constants that must be estimated in such a way that the MSE of the error is minimized.

By applying Holt-Winters method in production data of washing machine motor in 2017, forecast will be as following. ($\alpha = 0.9, \beta = 0.9, \gamma = 0.9$, number of periods=3, forecast range=3)

HOLT WINTERS

Motor Krkt.	OCAK	ŞUBAT	MART	NİSAN	MAYIS	HAZİRAN	TEMMUZ
0		48874	65193	44381	44792	31460	23112
ASKOL		0	0	0	0	0	0
ATLAS 22mm		12666	18713	15704	18154	12613	13254
ATLAS 32mm		86273	121031	126834	108230	67940	765506
ATLAS 42mm		22050	26562	28502	30973	23204	18752
ATLAS 50mm		325	304	29	359	515	8587
DIRECT DRIVE		1352	1481	613	787	814	2051
UNIVERSAL		163990	209126	206758	183059	137366	140016
Total		335656	443030	422771	388186	275133	274457

g. Autoregressive Model

Autoregressive (AR) model is a representation of a type of random process and it is used to describe certain time-varying processes. It specifies that the output variable depends linearly on its own previous values and on an imperfectly predictable term. The AR model of a random process in discrete time is defined by the following equation:

$y(t) = \sum_{i=1}^{p} a(i) * y(t-i) + \in(t)$ where $a_1, a_2 \ldots, a_p$ coefficients of the recursive filter; p is the order of the model; $\in(t)$ are output uncorrected errors.

h. ARIMA Model

ARIMA is a generalization of an autoregressive moving average (ARMA) model. Both models are fitted to time series data to predict future points in the series (forecasting). ARIMA models are applied in some cases where data show evidence of non-stationarity.

Autoregressive AR process: Series current values depend on its own previous values and AR(p) – Current values depend on its own p-previous values. p is the order of AR process.

Moving Average MA process: The current deviation from mean depends on previous deviations and MA(q) – The current deviation from mean depends on q- previous deviations. q is the order of MA process.

The full model can be written as: $\acute{y}_t = c + \emptyset_1 y_{t-1} + \cdots + \emptyset_p \acute{y}_{t-p} + \acute{\emptyset}_1 e_{t-1} + \cdots + \emptyset_q e_{t-q} + e_t$

where $\acute{y}_t$ is the differenced series. The predictors on the right hand side include both lagged values of y_t and lagged errors, this is called ARIMA(p,d,q) model. Where,

p= order of the autoregressive part;

d= degree of first differencing involved;

q= order of the moving average part.

The stationarity and invertibility conditions that are used for autoregressive and moving average models apply to this ARIMA model. Many of the ARIMA models have special cases as shown in the following.

White noise ARIMA(0,0,0): The errors are uncorrelated across time. This doesn't imply anything about the size of the errors, so in general it is not an indication of good or bad fit.

Random walk ARIMA(0, 1, 0) with a constant: If the series *y* is not stationary, the simplest possible model for it is a random walk model, which can be considered as a limiting case of an AR(1) model in which the autoregressive coefficient is equal to 1, i.e., a series with infinitely slow mean reversion.

Random walk ARIMA(0, 1, 0) with no constant: If the constant term in the random walk model is *zero*, it is a random walk without drift.

Autoregression ARIMA(*p*, 0, 0): If the series is stationary and autocorrelated, perhaps it can be predicted as a multiple of its own previous value, plus a constant.

Moving Average ARIMA(0, 0, *q*)

By applying ARIMA method in production data of washing machine motor in 2017, forecast will be as following. (Testing period=4, lambda=1, d=0, D=0, s=1, AR(p)=0, MA(q)=0)

	Motor Krkt.	OCAK	ŞUBAT	MART	NİSAN	MAYIS	HAZİRAN	TEMMUZ			
			Y[t]						F[t]	95% LB	95% UB
ARIMA	0		43503	66146	41160	41765	30398	22025	0.29	-119057	119057
	ASKOL		0	533	0	0	11	0	0	-648	648
	ATLAS 22mm		13160	21817	14517	19381	13687	12892	0	-31801	31801
	ATLAS 32mm		84991	117546	124887	103622	63177	80596	0	-188592	188592
	ATLAS 42mm		21241	23819	28736	29998	19704	19175	0	-49021	49021
	ATLAS 50mm		393	439	0	511	728	882	0	-713	713
	DIRECT DRIVE		1127	2096	495	743	1142	1801	0	-3185	3185
	UNIVERSAL		144850	236981	202088	158989	152312	142183	0	-356903	356903
	Total		309265	469377	411883	355010	281158	279553	0	-745673	745673

B. Associative Models

Linear Regression

Linear regression is a statistical technique for quantifying the relationship between variables (dependent variable and independent variables) and forecasting linear trend line. For example, assuming that the amount of advertising dollars spent on a product determines the amount of its sales, regression analysis can be used to quantify the precise nature of the relationship between advertising and sales.

Note: Dependent variable is the variable that one wants to forecast. Independent variables that are assumed to affect the dependent variable and thereby “cause” the results observed in the past.

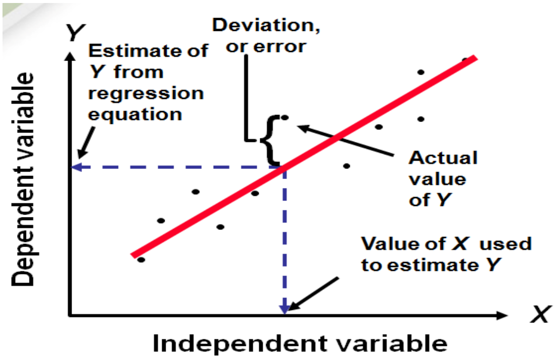
Regression equation: $\boldsymbol{Y = a + bX}$

Y = dependent variable

X = independent variable

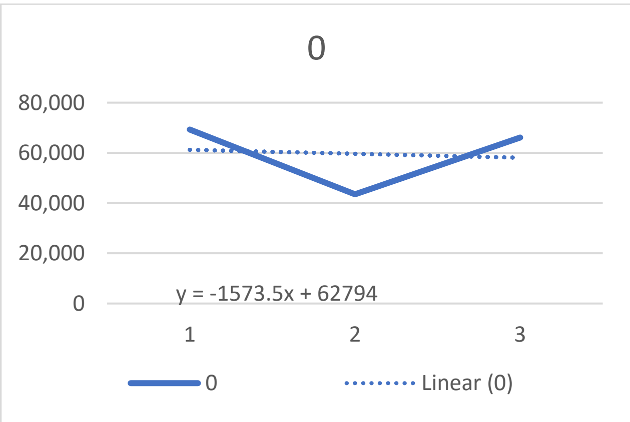
a = Y-intercept of the line

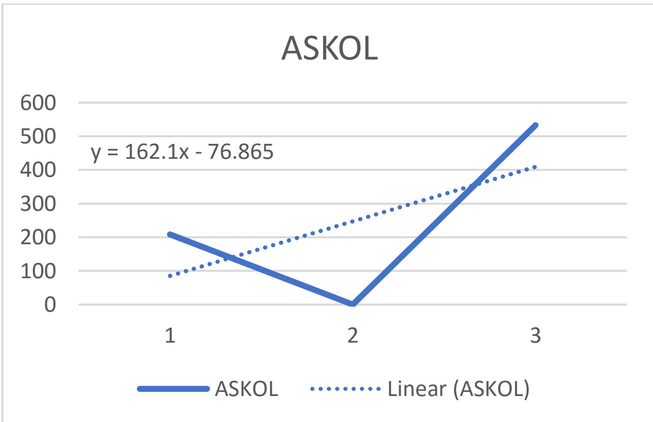
b = slope of the line

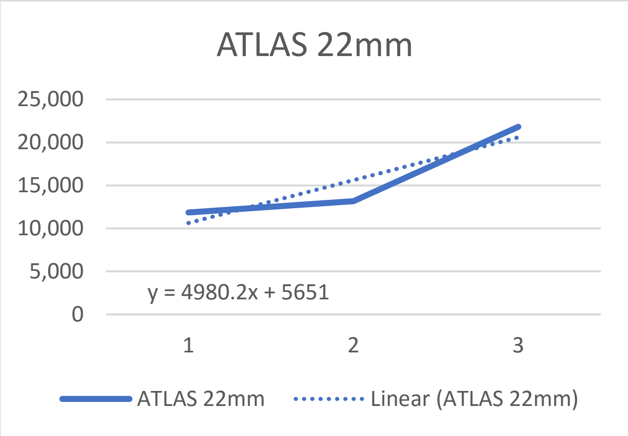


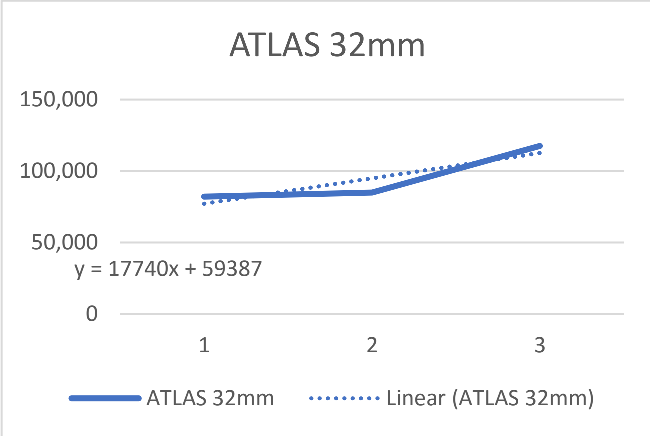
By applying linear regression method in production data of washing machine motor in 2017, forecast will be as following.

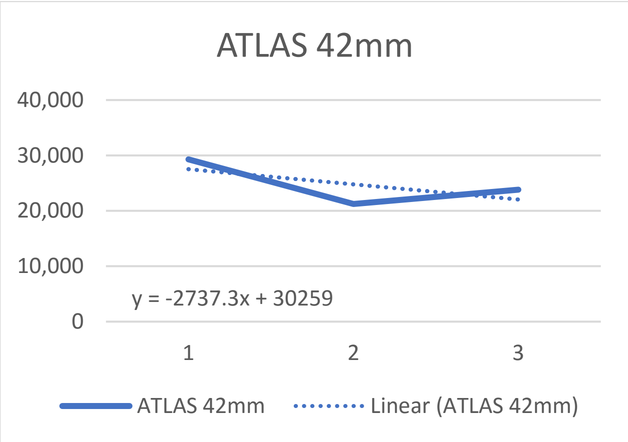
	Motor Krkt.	1	2	3	4	5	6	7
LINEAR REGRESSION	0				56500	54926.5	53353	51779.5
	ASKOL				-76216.6	-76054.5	-75892.4	-75730.3
	ATLAS 22mm				25571.8	24901	29881.2	34861.4
	ATLAS 32mm				130347	148087	165827	183567
	ATLAS 42mm				19309.8	16572.5	13835.2	11097.9
	ATLAS 50mm				567.91	675.75	783.59	891.43
	DIRECT DRIVE				2168.04	2464.42	2760.8	3057.18
	UNIVERSA L				264631	308401	352171	395941
	Total				499668	562414	625160	687906

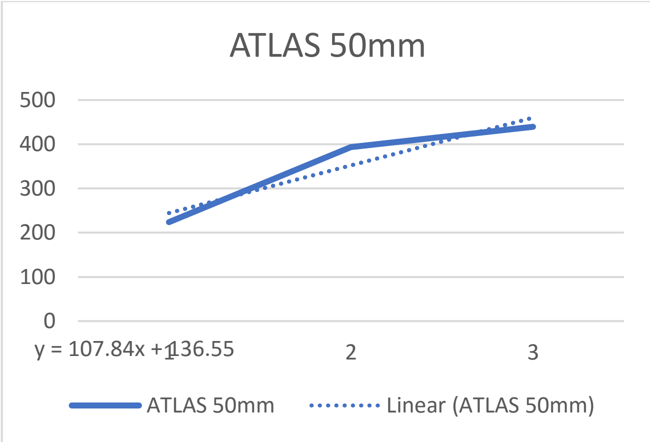


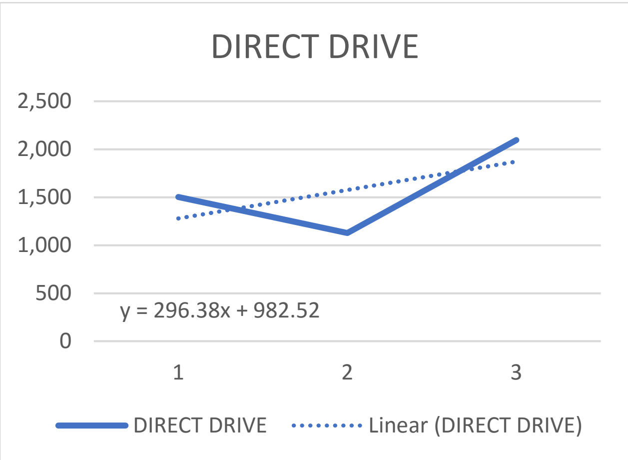


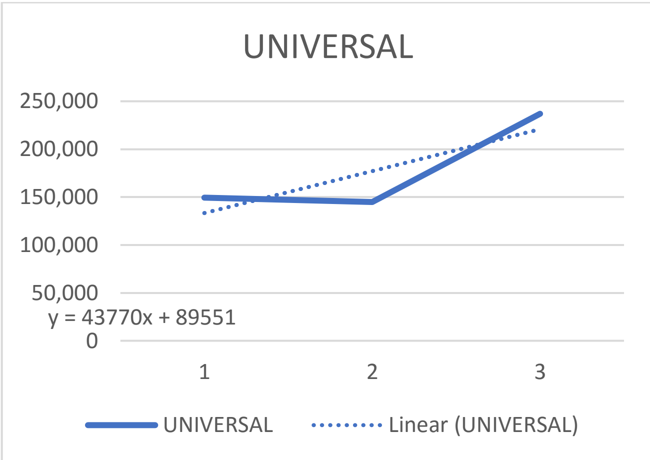












What is adjusted R^2?

Adjusted R^2 is a special form of coefficient of determination (R^2) and it indicates how data points fit a curve or line, but adjusts for the number of terms in a model. If we add more and more useless variables to a model, adjusted R^2 will decrease. If we add more useful variables, adjusted R^2 will increase. Adjusted R^2 will always be less than or equal to R^2. The formula is:

$$\boldsymbol{R^2_{adj} = 1 - \left[\frac{(1-R^2)*(n-1)}{n-k-1}\right]}$$

n= the number of points in data sample

k= the number of independent variables

1.2 Defining Forecast Accuracy

Forecast error: Difference between the actual value and the value that was predicted for a given period.

$\mathbf{E_t = D_t - F_t}$ where E_t = forecast error for period t, D_t = actual demand for period t, F_t = forecast period t.

Forecast Error Measures

- Cumulative sum of Forecast Errors (CFE)**:** A measurement of the total forecast error that assesses the bias in a forecast.

$$CFE = \sum E_t$$

- Mean Absolute Deviation (MAD): indicates on an average basis, how many units the forecast is off from the actual data.
  $\boldsymbol{MAD = \frac{\sum|E_t|}{n}}$ where $|E_t|$ = absolute value of the forecast error for period t, number of periods of evaluation.

- Mean Absolute Percent Error (MAPE): indicates on an average basis, how many percent the forecast is off from the actual data.

$$\boldsymbol{MAPE = \frac{\sum\left[\frac{|E_t|}{D_t}\right]*100}{n}}$$ where E_t = forecast error for period t, D_t = actual demand for period t.

- Mean Squared Error (MSE): A forecast error measure that penalises large errors proportionally more than small errors.

$$\boldsymbol{MSE = \frac{\sum E_t^2}{n}}$$

- Tracking Signal: A measure that indicates whether a method of forecasting is accurately predicting actual changes in demand.

$$Tracking\ signal = \frac{CFE}{MAD}$$

When forecasting accuracy is defined for forecasting methods which have been previously applied on production numbers of Arçelik for washing machine motor in 2017, the result will be as following.

NAIVE

Motor Krkt.	CFE	TS	MAD	MSE	MAPE
0	-47,268	-3	15627	333637353	7%
ASKOL	-209	-1	216	102038	63348%
ATLAS 22mm	1,035	0	4769	31109106	1894%
ATLAS 32mm	-1,469	0	20325	585620396	70%
ATLAS 42mm	-10,119	-2	4606	33920232	15%
ATLAS 50mm	658	3	256	92689	2229%
DIRECT DRIVE	298	0	709	716062	35834%
UNIVERSAL	-7,258	0	31920	1955243828	540%
Total	-64,332	-1	64093	6471903819	65%

SIMPLE MOVING AVERAGE

Motor Krkt.	CFE	TS	MAD	MSE	MAPE
0	-62,034	-4	15508	258585952	1146%
ASKOL	-595	-4	149	30149	137188%
ATLAS 22mm	-6,066	-2	2958	10548020	489%
ATLAS 32mm	-44,307	-2	26087	982622567	701%
ATLAS 42mm	-5,434	-1	6034	38606153	618%
ATLAS 50mm	761	2	366	141857	1727%
DIRECT DRIVE	-538	-1	654	607378	1563%
UNIVERSAL	-86,641	-3	34159	1236642040	521%
Total	-204,854	-3	70067	6296639768	528%

WEIGHTED MOVING AVERAGE

Motor Krkt.	CFE	TS	MAD	MSE	MAPE
0	-55,261	-4	13815	203794923	1021%
ASKOL	-562	-4	141	32126	129583%
ATLAS 22mm	-7,408	-2	3180	10681302	526%
ATLAS 32mm	-44,264	-2	23119	813728578	621%
ATLAS 42mm	-5,296	-1	5831	37242482	597%
ATLAS 50mm	658	2	358	130069	1690%
DIRECT DRIVE	-413	-1	681	604555	1630%
UNIVERSAL	-89,476	-3	27573	933943274	421%
Total	-202,024	-3	58902	4643820343	444%

EXPONENTIAL SMOOTHING

Motor Krkt.	CFE	TS	MAD	MSE	MAPE
0	-51,462	-3	15265	303745173	486%
ASKOL	-231	-1	211	90106	2806%
ATLAS 22mm	1,297	0	4555	27920286	424%
ATLAS 32mm	-3,096	0	20373	597566609	310%
ATLAS 42mm	-11,072	-2	4725	33636028	275%
ATLAS 50mm	711	3	263	90224	827%
DIRECT DRIVE	254	0	670	657825	752%
UNIVERSAL	-6,814	0	31693	1904270781	267%
Total	-70,414	-1	63952	6283473683	261%

LINEAR REGRESSION

Motor Krkt.	CFE	TS	MAD	MSE	MAPE
0	-81,211	-4	20302.805	455204937.2	1500%
ASKOL	303,905	4	75976.163	5772409756	70013869%
ATLAS 22mm	-54,740	-4	13684.876	224405163.4	2263%
ATLAS 32mm	-255,547	-4	63886.664	5786770231	1716%
ATLAS 42mm	36,797	4	9199.3356	92193987.58	942%
ATLAS 50mm	-797	-4	199.29866	88169.56942	939%
DIRECT DRIVE	-6,269	-4	1567.1274	2489505.698	3747%
UNIVERSAL	-665,572	-4	166393.12	32643119330	2538%
Total	-1,047,544	-4	261886.11	83953073709	1973%

HOLT WINTERS

Motor Krkt.	CFE	TS	MAD	MSE	MAPE
0	-12,816	-5.22329	2,454	8,601,036	78%
ASKOL	544	6	91	47,383	1204%
ATLAS 22mm	4,349	3.503487	1,241	2,346,446	116%
ATLAS 32mm	-700,996	-6	116,833	78,193,847,194	1779%
ATLAS 42mm	-7,370	-5.09203	1,447	3,602,335	84%
ATLAS 50mm	-7,165	-5.1768	1,384	9,910,736	4355%
DIRECT DRIVE	307	1.166122	263	102,366	295%
UNIVERSAL	-2,913	-0.18823	15,475	328,579,827	130%
Total	-32,987	-1.83392	17,987	445,360,495	73%

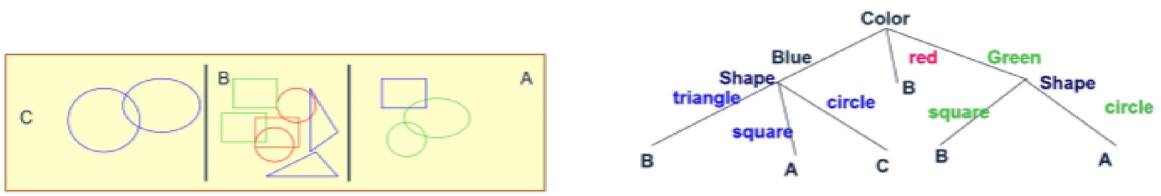
ARIMA	Motor Krkt.	MSE	MAPE
	0	1096324196	1
	ASKOL	17.64	0
	ATLAS 22mm	196514846	1
	ATLAS 32mm	8120252926	1
	ATLAS 42mm	723116803	1
	ATLAS 50mm	196630	0
	DIRECT DRIVE	608790	1
	UNIVERSAL	22553384127	1
	Total	96492386696	1
		14353642781	78%

FORECASTING METHODS	CFE	TS	MAD	MSE	MAPE
NAÏVE	-14,296	0	15,836	1,045,816,169	116
SIMPLE MOVING AVERAGE	-45,523	-2	17,331	980,491,543	161
WEIGHTED MOVING AVERAGE	-44,894	-2	14,845	738,219,739	15170%
EXPONENTIAL SMOOTHING	-15,647	-1	15,745	1,016,827,857	712%
LINEAR REGRESSION	-196,775	-2	68,122	14,325,528,310	7781054%
HOLT-WINTERS	-84,339	-1	17,464	8,776,933,091	902%
ARIMA				14353642781	78%

2. DECISION TREE

A decision tree is a decision support tool that uses a model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal.

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with **decision nodes** and **leaf nodes**. In classification, the goal is to learn a decision tree that represents the training data such that labels for new examples can be determined. Decision trees are classifiers for instances represented as feature vectors (e.g. color=?; shape=?; label=?). Nodes are tests for feature values, leaves specify the label, and at each node there must be one branch for each value of the feature. A decision node (e.g., color) has two or more branches (e.g., shape). Leaf node (e.g., label) represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called **root node**. Decision trees can handle both categorical and numerical data.



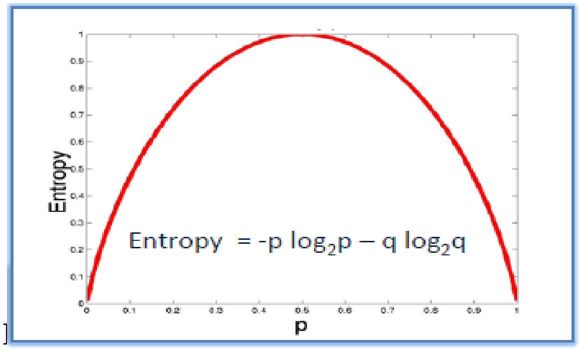
(a) Example Data (b) Decision Tree

Figure 1: Decision Tree Example

From the example in Figure 1, given a new shape, we can use the decision tree to predict its label.

2.1 Algorithm of decision tree: The core algorithm for building decision trees called **ID3** by J. R. Quinlan which employs a top-down, greedy search through the space of possible branches with no backtracking. ID3 uses *Entropy* and *Information Gain* to construct a decision tree.

Entropy: A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous). ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has entropy of one.

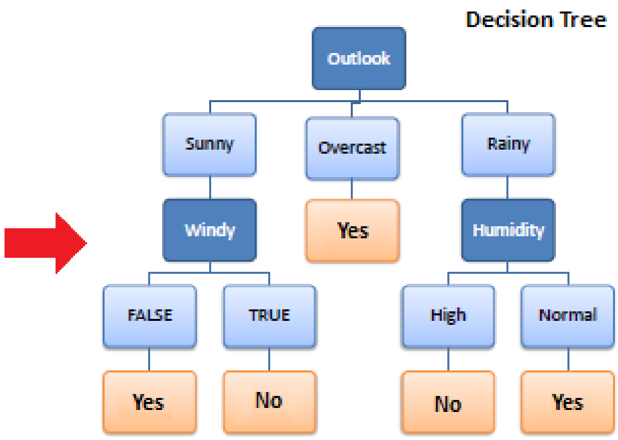


Information Gain: The information gain is based on the decrease in entropy after a dataset is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain (i.e., the most homogeneous branches).

Entropy = -0.5 $\log_2$0.5 – 0.5 $\log_2$0.5 = 1

rget

Outlook	Temp.	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



To build a decision tree, firstly we need to calculate two types of entropy using frequency tables as follows:

a) Entropy using the frequency table of one attribute:

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Play Golf	
Yes	No
9	5

Entropy(PlayGolf) = Entropy (5,9)
= Entropy (0.36, 0.64)
= - (0.36 $\log_2$ 0.36) - (0.64 $\log_2$ 0.64)
= 0.94

b) Entropy using the frequency table of two attributes:

$$E(T,X) = \sum_{c \in X} P(c)E(c)$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

E(PlayGolf, Outlook) = **P**(Sunny)***E**(3,2) + **P**(Overcast)***E**(4,0) + **P**(Rainy)***E**(2,3)

= (5/14)*0.971 + (4/14)*0.0 + (5/14)*0.971

= 0.693

2.2 How to draw a decision tree?

Step 1: Calculate entropy of the target.

Entropy(PlayGolf) = Entropy (5,9)

= Entropy (0.36, 0.64)

= - (0.36 $\log_2$ 0.36) - (0.64 $\log_2$ 0.64)

= 0.94

Step 2: The dataset is then split on the different attributes. The entropy for each branch is calculated. Then it is added proportionally, to get total entropy for the split. The resulting entropy is subtracted from the entropy before the split. The result is the Information Gain, or decrease in entropy.

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1
Gain = 0.029			

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1
Gain = 0.152			

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3
Gain = 0.048			

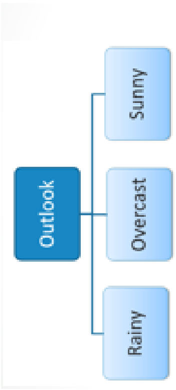
$$Gain(T,X) = Entropy(T) - Entropy(T,X)$$

G(PlayGolf, Outlook) = **E**(PlayGolf) – **E**(PlayGolf, Outlook)

= 0.940 – 0.693 = 0.247

Step 3: Choose attribute with the largest information gain as the decision node, divide the dataset by its branches and repeat the same process on every branch.

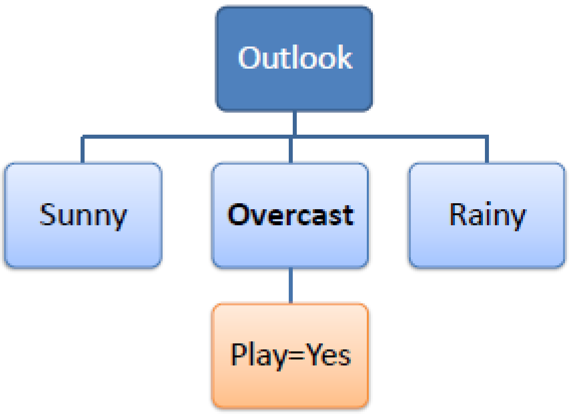
★		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			



Outlook	Temp	Humidity	Windy	Play Golf
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Sunny	Mild	Normal	FALSE	Yes
Sunny	Mild	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
Overcast	Cool	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes

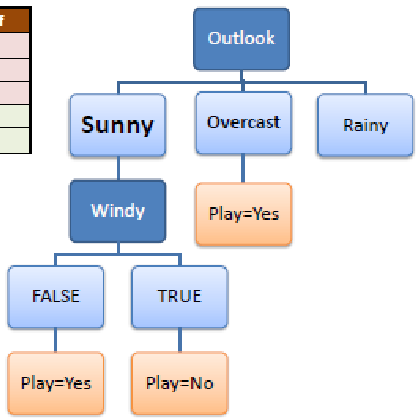
Step 4a: A branch with entropy of 0 is a leaf node.

Temp	Humidity	Windy	Play Golf
Hot	High	FALSE	Yes
Cool	Normal	TRUE	Yes
Mild	High	TRUE	Yes
Hot	Normal	FALSE	Yes



Step 4b: A branch with entropy more than 0 needs further splitting.

Temp	Humidity	Windy	Play Golf
Mild	High	FALSE	Yes
Cool	Normal	FALSE	Yes
Mild	Normal	FALSE	Yes
Cool	Normal	TRUE	No
Mild	High	TRUE	No



Step 5: The ID3 algorithm is run recursively on the non-leaf branches, until all data is classified.

2.3 Decision Tree Methods

i) CHAID (Chi-squared Automatic Interaction Detector)

CHAID method is based on the chi-square test of association. A CHAID tree is a decision tree that is constructed by repeatedly splitting subsets of the space into two or more child nodes, beginning with the entire data set.

ii) QUEST (Quick, Unbiased, Efficient Statistical Tree)

QUEST is a binary-split decision tree algorithm for classification and data mining. QUEST can be used with univariate or linear combination splits.

iii) CART (Classification and Regression Tree)

CART is a recursive partitioning method to be used both for regression and classification. CART is constructed by splitting subsets of the data set using all predictor variables to create two child nodes repeatedly, beginning with the entire data set.

Works Cited

"Adjusted R2 / Adjusted R-Squared: What is it used for?" Statistics How To. N.p., n.d. Web. 13 Aug. 2017.

"ARIMA models for time series forecasting." People.duke.edu, 2017. Web. https://people.duke.edu/~rnau/411arim.htm Accessed 13 Aug. 2017.

"Autoregressive model." Wikipedia. Wikimedia Foundation, 13 Aug. 2017. Web. 13 Aug. 2017.

"Dangers in Using Only Trend Analysis in Forecasting." *Chron.com*, smallbusiness.chron.com/dangers-using-only-trend-analysis-forecasting-37447.html. Accessed 16 Aug. 2017.

"Decision tree." Wikipedia. Wikimedia Foundation, 09 Aug. 2017. Web. 13 Aug. 2017.

"Decision tree." Saedsayad.com, 2017. Web. http://www.saedsayad.com/decision_tree.htm Accessed 16 Aug. 2017.

"Forecasting Techniques." LinkedIn SlideShare. N.p., 06 Dec. 2009. Web. 13 Aug. 2017.

J. Hueter and W. Swart "An Integrated Labor-Management System for Taco Bell," *Interfaces*, 28, no. 1 (January-February 1998): 75-91.

Knksmart Follow. "Forecasting Slides." LinkedIn SlideShare. N.p., 17 June 2008. Web. 13 Aug. 2017.

Trubetskoy, Gregory. "Holt-Winters Forecasting for Dummies." Holt-Winters Forecasting for Dummies (or Developers) - Part I - GrishaTrubetskoy. N.p., n.d. Web. 13 Aug. 2017.

Zhou, B., & Cervantes, C. (2016, September 8). *Decision Trees* [PDF]. University of Illinois Urbana-Champaign.