SAS® GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

Forecasting Frequency of Terrorist attacks in United States and Predicting the Factors that determine Success of Attacks using SAS Forecast Studio™ and SAS Enterprise Miner™

Sri Harsha Ramineni
Koteswarrao Sriram

#SASGF

Forecasting Frequency of Terrorist attacks in United States and Predicting the Factors that determine Success of Attacks using SAS Forecast Studio™ and SAS Enterprise Miner™

SriHarsha Ramineni, Koteswara Rao Sriram

MS in MIS, SAS® and OSU Data Mining Certificate, Oklahoma State University

ABSTRACT

The importance of understanding terrorism in the United States of America assumed heightened prominence in the wake of the coordinated attacks of September 11, 2001. Yet surprisingly little is known about the frequency of attacks that may happen in the future and the factors that lead to a successful terrorist attack. This research is aimed at forecasting the frequency of attacks per annual basis in United States and to determine the factors that contribute to its success.

Using the data acquired from Global Terrorism Database (GTD), we tested our hypothesis by examining the frequency of attacks on annual basis from 1972 to 2014 using SAS® Enterprise Miner and SAS® Forecasting Studio. The data set has 2,683 observations. Our Forecasting model predicts that there may be at least 16 attacks every year for the next 4 years. From our study of factors which contribute to the success of the terrorist attack, we discovered that attack type, weapons used in the attack, place of attack and type of target play pivotal roles in determining the success of a terrorist attack. Results reveal that the government may be successful in averting assassination attempts but may fail to prevent armed assaults and facilities/infrastructure attacks. So additional security may need to be provided to important facilities to prevent further attacks from being successful. Also results further reveal that it is possible to reduce the forecasted number of attacks by raising defense spending and by putting an end to the raging war in middle east.

METHOD

• Majority of the data was gathered from Global Terrorism Database (GTD) whereas the rest of the data was collected from various sources which include U.S Economy, U.S census and so on.

• Two different data sets were used for forecasting the frequency and predicting the outcome of a terrorist attack. For forecasting, the dataset with 11 variables was used with success as target whereas the dataset used for forecasting has 8 variables.

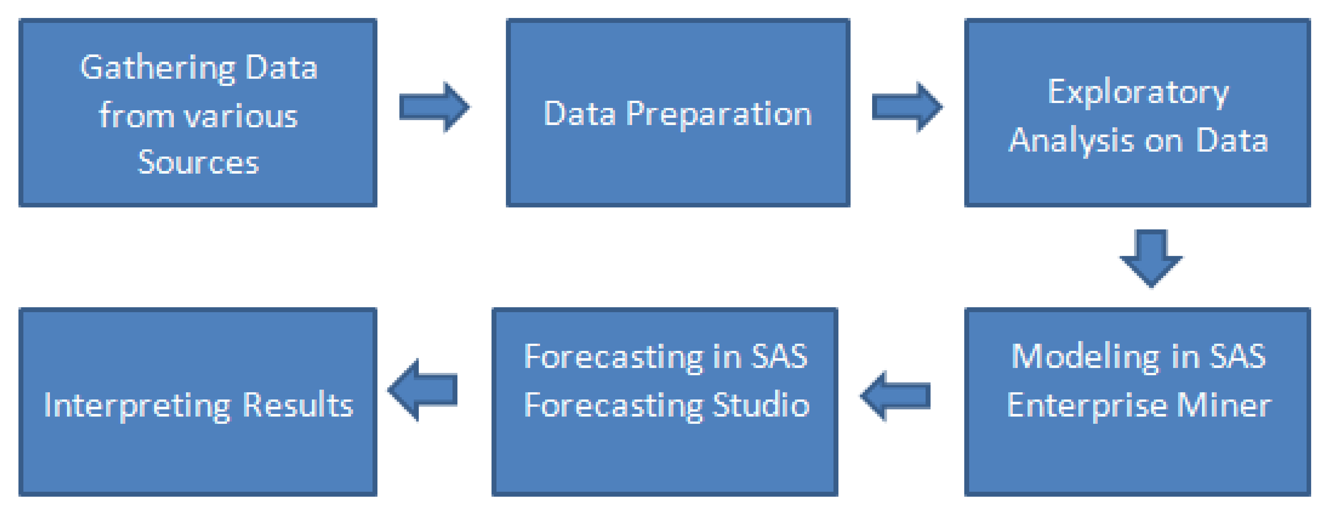


Figure 1: Procedure followed for Research

•For missing values, mode imputation for class variables and mean imputation for interval variables was performed.

•The data set that had information about terrorist attacks in different countries was then filtered to obtain terrorist attacks in U.S. alone.

RESULTS

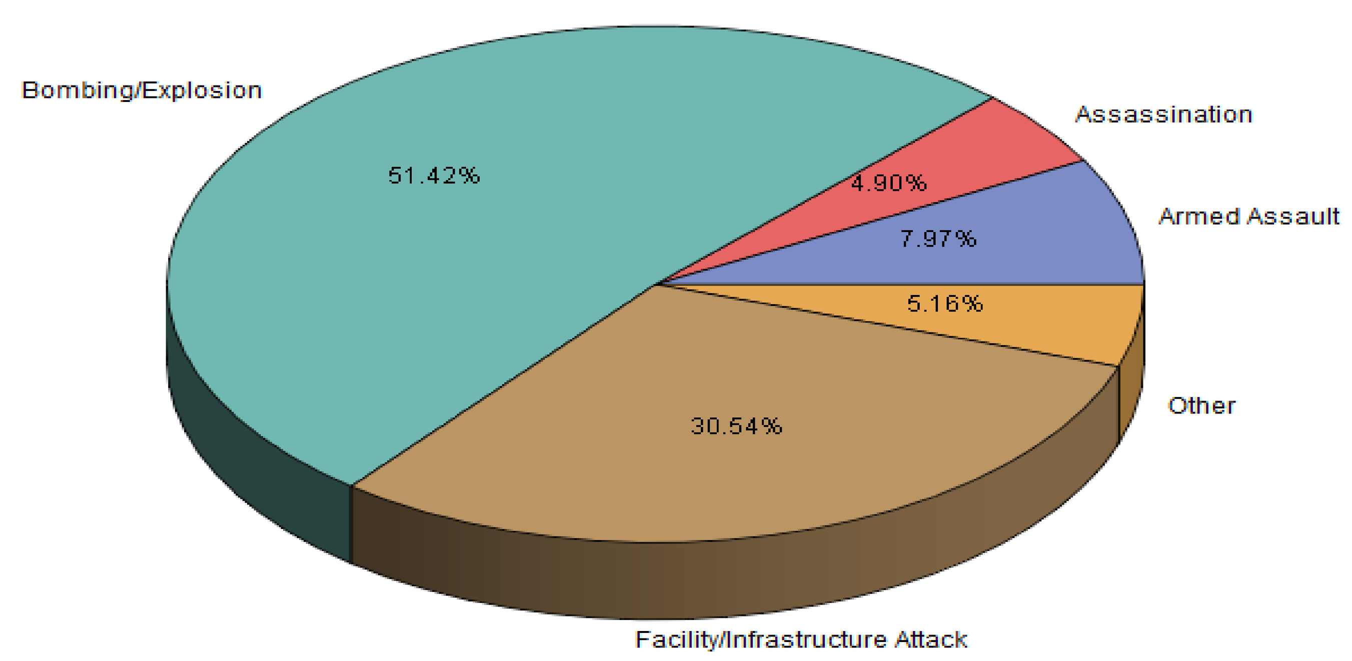


Figure 2: Type of attack

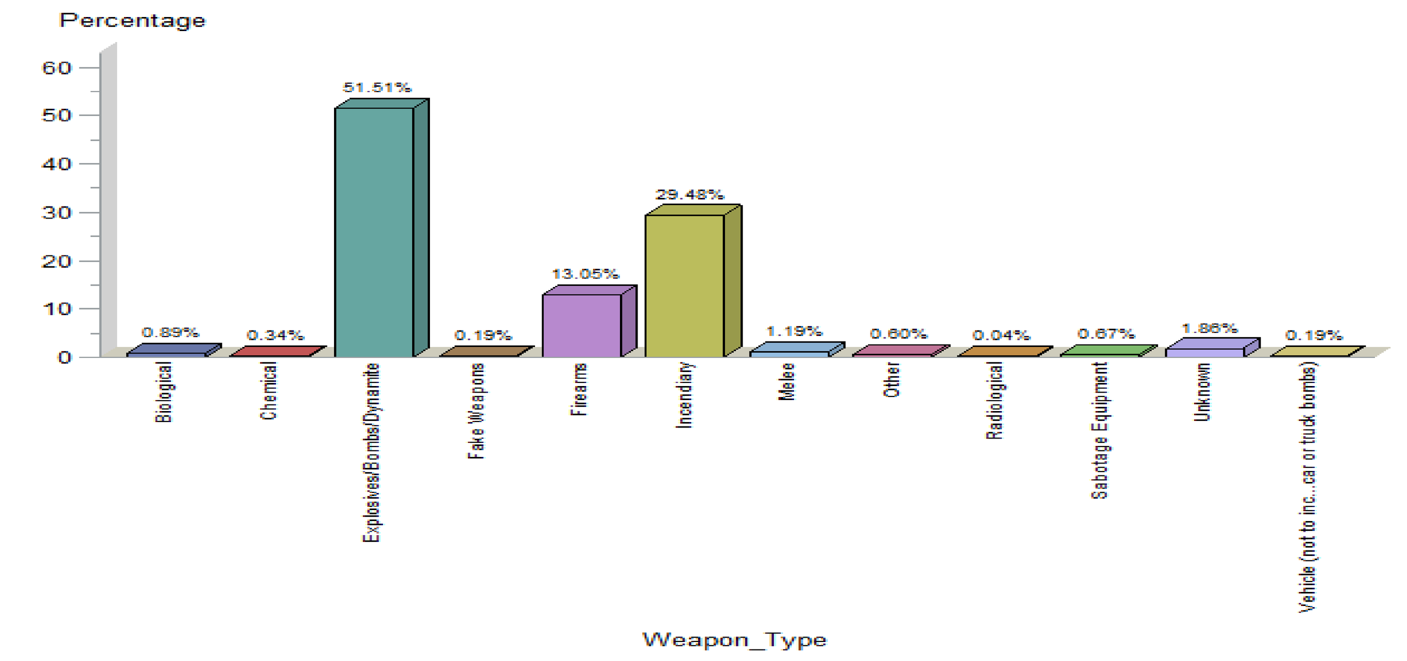


Figure 3: Weapon Type

Exploratory Analysis:

•Out of all the events, Bombing/Explosion was used on many instances which contributes around 51.42%, next highest in the list is Facility/Infrastructure attack with 30.54% and third one in the list is Armed Assault with 7.97% of all the terrorist events happened until 2014.

•. Weapon type of Explosives/Bombs/Dynamite was most widely used by terrorists with 51.51%, next in the list is Incendiary with 29.48% and third one in the list is Firearms with 13.05% whereas all other weapon type's contribution is very low when compared to the top three types.

•An Interesting fact is that there is not even a single state without any terrorist attack in the last 44 years. Highest number of terrorist attacks happened in California (588), second in New York (500) and the lowest in Rhode Island, Kentucky and Wyoming (2). Coastal states are more prone to attacks.

•For target type, Business entities are in the top with 27.58%, followed by Government entities with 12.25% and third is Private Citizens & Property with 11.43% of all the terrorist events.

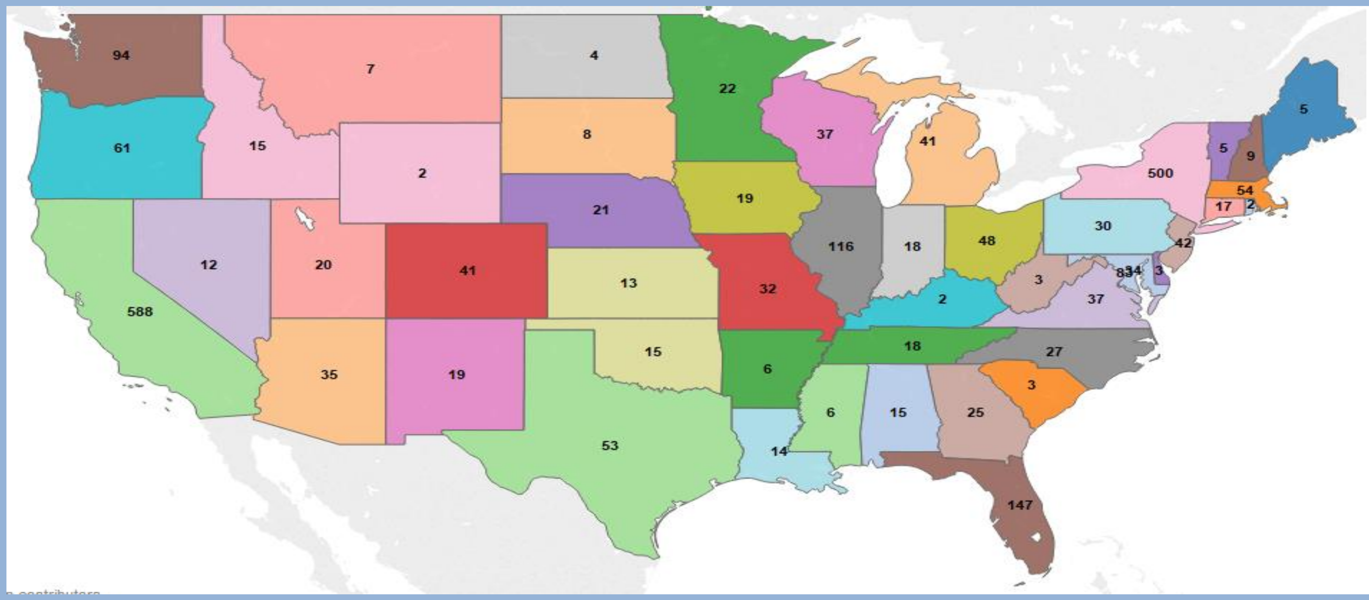


Figure 4: Attack Location

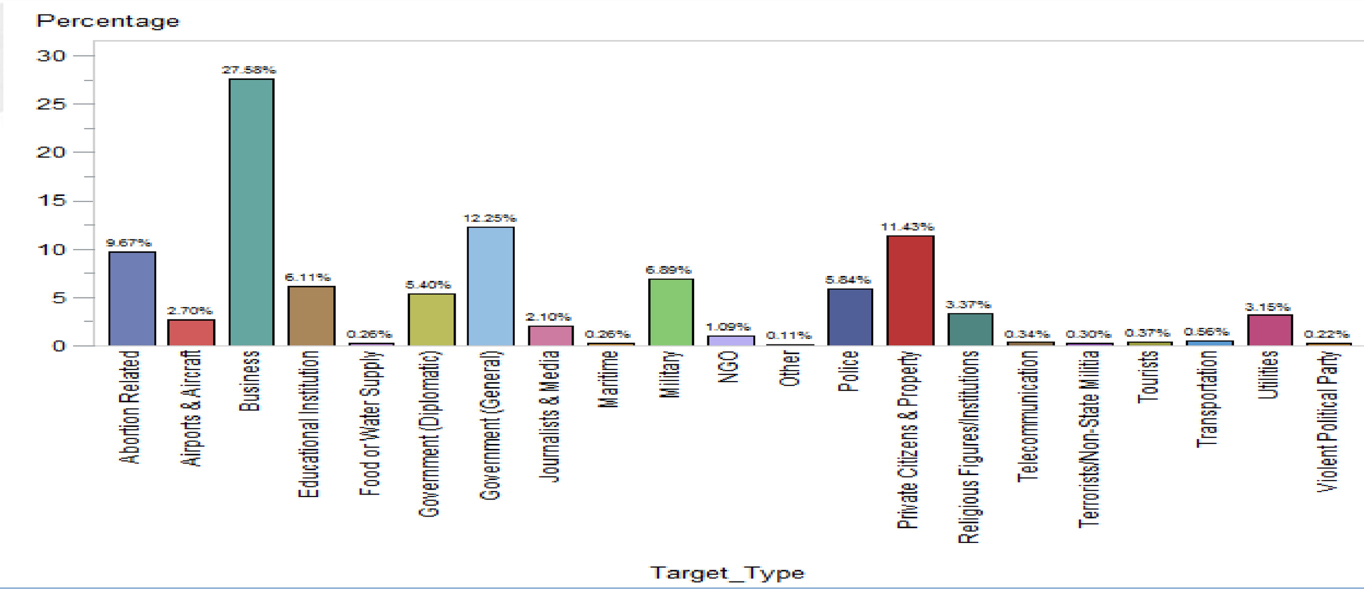


Figure 5: Target Type

Forecasting Frequency of Terrorist attacks in United States and Predicting the Factors that determine Success of Attacks using SAS Forecast Studio™ and SAS Enterprise Miner™

SriHarsha Ramineni, Koteswara Rao Sriram

MS in MIS, SAS® and OSU Data Mining Certificate, Oklahoma State University

RESULTS CONTINUED

•Frequency of attacks has a range of 143 with 149 being the highest and 6 being the lowest. The highest number of attacks occurred in the year of 1975 whereas the lowest number of attacks occurred in the year of 2006.

•Population in the U.S is steadily rising over the years with 2014 being the highest with a population of 318.86 million and 1972 being the least with 108.96 million.

•U.S defense budget had a significant rise over the years. This might be a reason which led to the less frequency of attacks in the recent years compared to the past.

•There is no particular trend in the GDP growth over the years and is following a random trend. The economy had a highest growth of 7.3% in the year 1984 and had a breakdown in the year 2009 with a value of -2.8%.

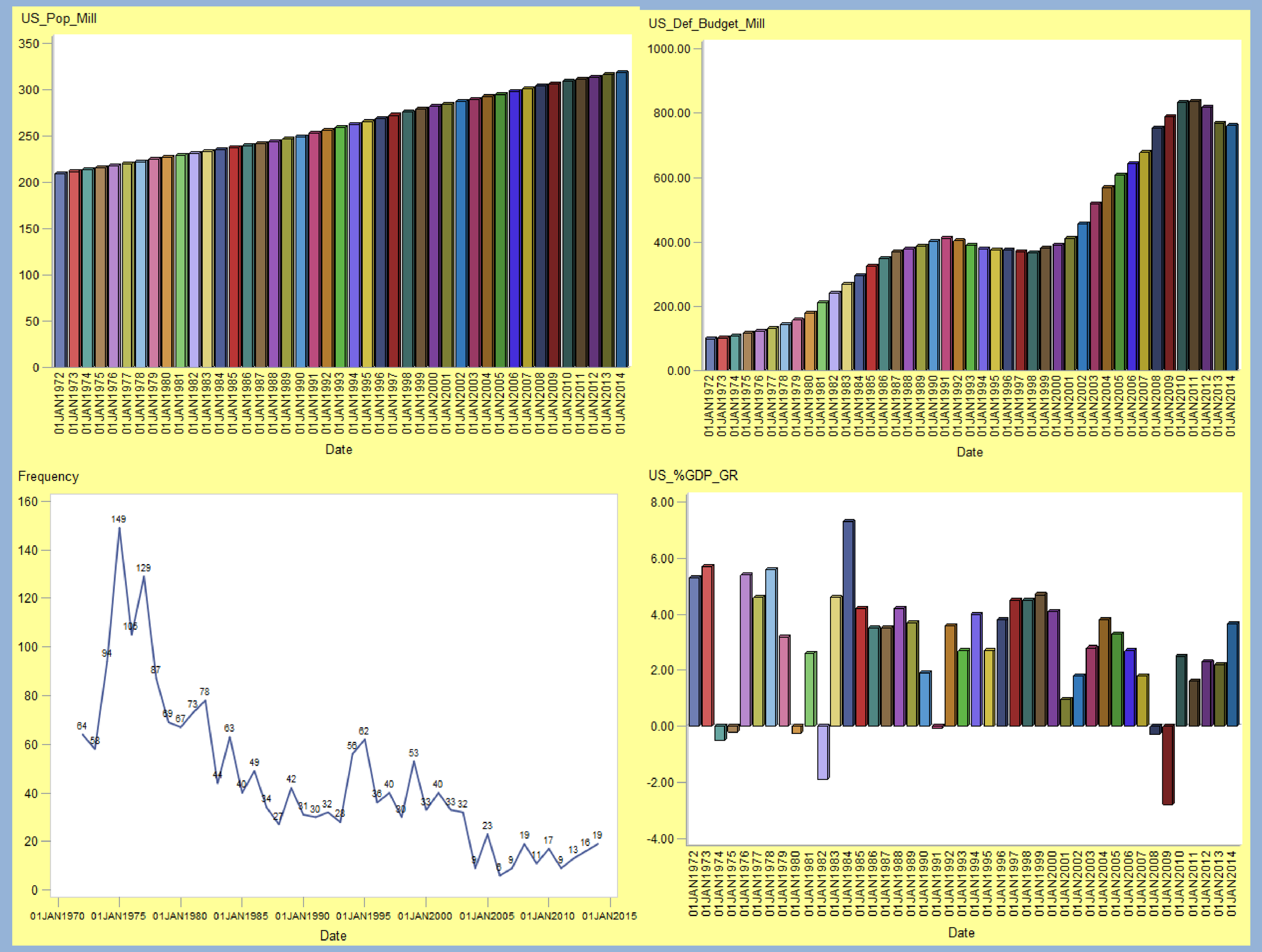


Figure 6: Exploratory Analysis of Independent variables

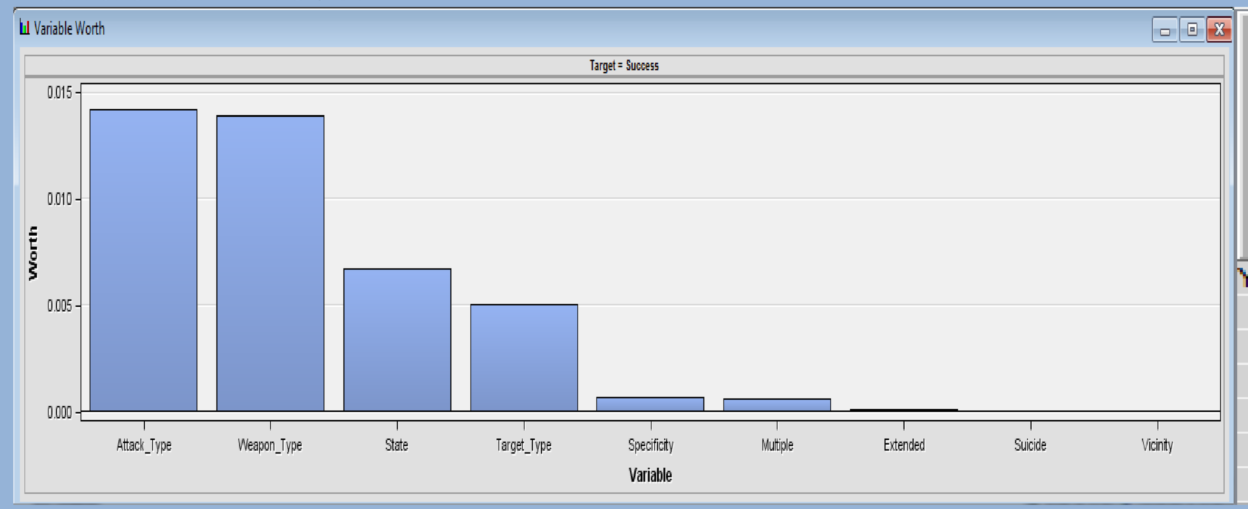


Figure 7: Variable Importance

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
Y	Neural	Neural	Neural Net...	Success		0.163569
	AutoNeural	AutoNeural	AutoNeural	Success		0.163569
	Ensmbl	Ensmbl	Ensemble	Success		0.164808
	Tree	Tree	Decision Tr...	Success		0.167286
	LARS	LARS	LARS	Success		0.17596
	Reg	Reg	Regression	Success		0.179678
	SVM	SVM	SVM	Success		0.180917

Figure 8: Model Comparison

Modeling using SAS Enterprise Miner™:

•Attack type & Weapon type are most important variables followed by State and Target type in predicting the success/failure of a terrorist attack. Other variables Specificity, Multiple, Extended, Suicide & Vicinity don't contribute much in predicting the success/failure of a terrorist attack.

•Models were built to predict the outcome of a terrorist attack with success as target and of all the model built Neural Network was the best model in terms of Validation Misclassification Rate of 0.163569.

RESULTS CONTINUED

•The neural network has a Validation Misclassification Rate of 0.162569 and Validation Average Squared Error of 0.134071.

Forecasting Frequency using SAS Forecast Studio™:

•Out of all the models built, ARIMA model with events turned out to be the best model with a MAPE value of 21.21.

•Population and Presidential Elections turned out to be significant independent variables.

•There is one significant outlier at 1977 with an estimate of 39.64 and a high t-value of 6.31.

Frequency, Frequency: MAPE= 21.21

Active series Forecast Model: LEAF_4COPY1(automatic selection) Set this model as forecast model

Model	Type	Read-Only	MAPE
LEAF_4COPY1	Custom	No	21.21
Generated ARIMA Model (LEAF_4)	Generated	Yes	21.57
Generated ARIMA Model (LEAF_5)	Generated	Yes	24.55
LEAF_4COPY2	Custom	No	25.40
COMBINEDMODEL1	Combination (Custom)	Yes	26.43
LEAF_5COPY1	Custom	No	32.93
Generated Smoothing Model (LEAF_6)	Generated	Yes	35.16

Figure 10: Model statistics for Forecasting

•*PULSE_1983*, *PULSE_1999*, *PULSE_2008*, *TEMP_1975_1978* and *LS_1994_1995* are significant at 5% level of significance. The event *TEMP_1983_1984* deems to be insignificant but still has an estimate of 21.37.

•All the prediction errors for frequency for forecasting lie within two standard deviation which indicate the model is significant.

•Also if we look at the ACF, PACF and IACF plots all the values lie within which indicates that the series is stationary.

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Success		_DFT_	Total Degrees of Freedom	1876	.	.
Success		_DFE_	Degrees of Freedom for Error	1584	.	.
Success		_DFM_	Model Degrees of Freedom	292	.	.
Success		_NW_	Number of Estimated Weights	292	.	.
Success		_AIC_	Akaike's Information Criterion	2114.591	.	.
Success		_SBC_	Schwarz's Bayesian Criterion	3731.365	.	.
Success		_ASE_	Average Squared Error	0.126007	0.134071	.
Success		_MAX_	Maximum Absolute Error	0.955415	0.973813	.
Success		_DIV_	Divisor for ASE	3752	1614	.
Success		_NOBS_	Sum of Frequencies	1876	807	.
Success		_RASE_	Root Average Squared Error	0.354974	0.366157	.
Success		_SSE_	Sum of Squared Errors	472.7768	216.3904	.
Success		_SUMW_	Sum of Case Weights Times Freq	3752	1614	.
Success		_FPE_	Final Prediction Error	0.172464	.	.
Success		_MSE_	Mean Squared Error	0.149235	0.134071	.
Success		_RFPE_	Root Final Prediction Error	0.415287	.	.
Success		_RMSE_	Root Mean Squared Error	0.38631	0.366157	.
Success		_AVERR_	Average Error Function	0.40794	0.432379	.
Success		_ERR_	Error Function	1530.591	697.8598	.
Success		_MISC_	Misclassification Rate	0.158849	0.163569	.
Success		_WRONG_	Number of Wrong Classifications	298	132	.

Figure 9: Best model fit statistics

Component	Parameter	Estimate	Standard Error	t Value	App Pr >
Frequency	AR1_1	-0.64938	0.16418	-3.96	0.000
US_Pop_Mill	SCALE	14.52878	6.28475	2.31	0.029
US_Pop_Mill	NUM1_1	14.69690	6.35521	2.31	0.029
US_Presidential_Elections	SCALE	-5.28430	2.40169	-2.20	0.037
US_Def_Budget_Mill	SCALE	-0.06325	0.03514	-1.80	0.084
US_GDP_PERCENT_GR	SCALE	-0.50713	0.70632	-0.72	0.479
PULSE_1983	SCALE	-33.17478	15.76197	-2.10	0.045
PULSE_1999	SCALE	15.71264	6.38714	2.46	0.021
PULSE_2008	SCALE	12.48140	6.80501	1.83	0.078
TEMP_1975_1978	SCALE	73.54095	18.38807	4.00	0.000
TEMP_1983_1984	SCALE	21.37472	15.52991	1.38	0.180
LS_1994_1995	SCALE	14.63408	5.73402	2.55	0.017
AO01JAN1977D	SCALE	39.64768	6.28639	6.31	<.000

Figure 11: Parameter Estimates of Best Model

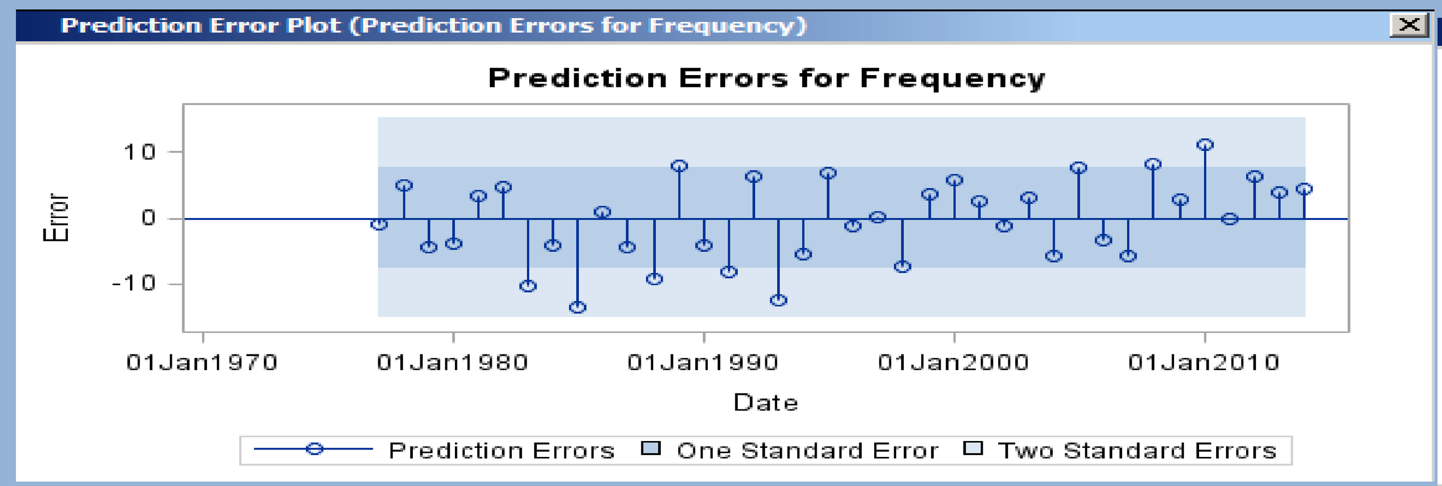


Figure 12: Prediction Errors Plot

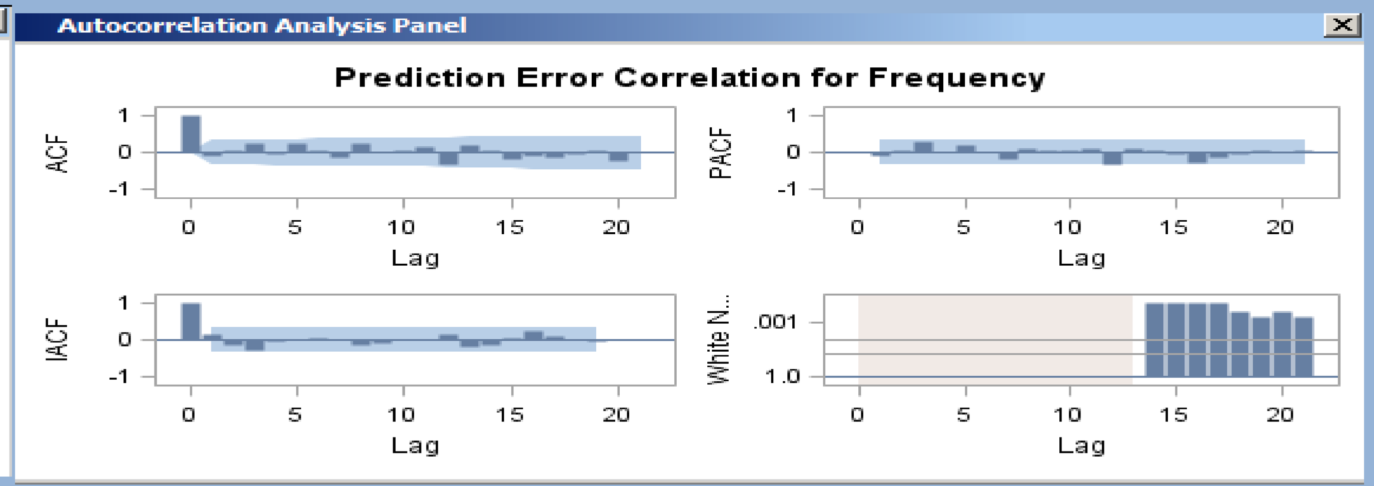


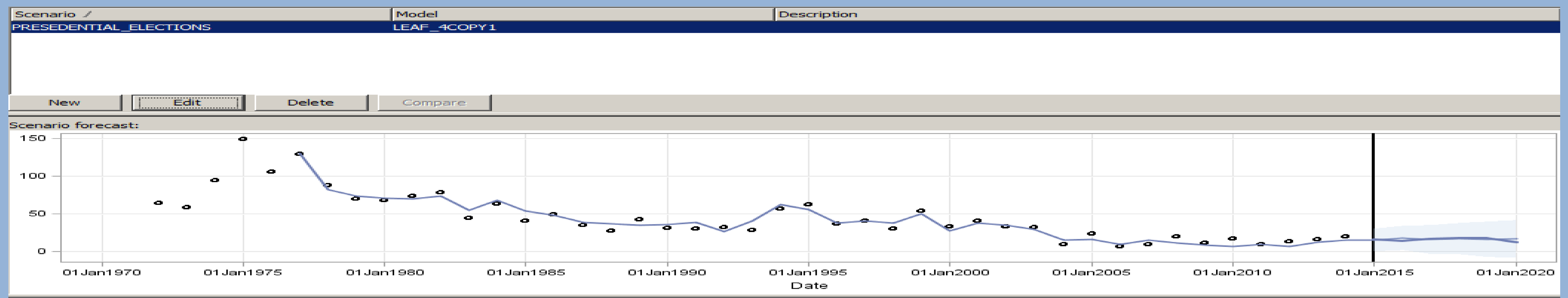
Figure 13: Prediction Error Correlation Plots

Forecasting Frequency of Terrorist attacks in United States and Predicting the Factors that determine Success of Attacks using SAS Forecast Studio™ and SAS Enterprise Miner™

SriHarsha Ramineni, Koteswara Rao Sriram

MS in MIS, SAS® and OSU Data Mining Certificate, Oklahoma State University

RESULTS CONTINUED

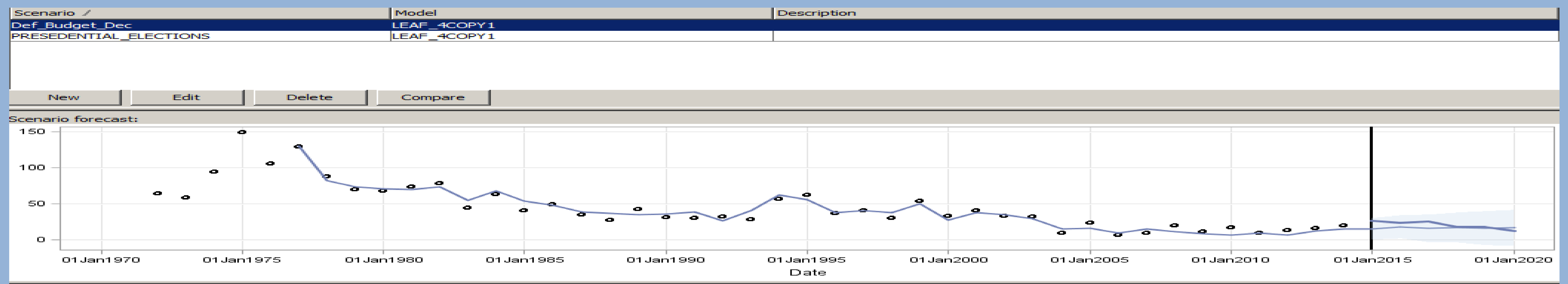


	01Jan2013	01Jan2014	01Jan2015	01Jan2016	01Jan2017	01Jan2018	01Jan2019	01Jan2020
* US_POP_MILL	316.5	318.86	321.22009029	323.58015057	325.94021085	328.30027112	330.6603314	333.02039167
* US_PRESIDENTIAL_ELECTIONS	0	0	0	1	0	0	0	1
* US_DEF_BUDGET_MILL	769.93	761.5	751.15806487	743.33944907	736.69184921	731.03987895	726.23442008	722.14868821
* US_GDP_PERCENT_GR	2.2	3.66	2.0965668093	2.0597967778	2.0230267464	1.986256715	1.9494866835	1.9127166521
Model Forecast	980123652	14.562493545	14.804013501	17.737307355	15.506210877	16.676862383	15.829475105	16.211564524
Scenario Forecast	.	.	15.993458588	13.642462806	16.695683018	17.866347205	17.018972076	12.116769679

Figure 14: Scenario 1

Scenarios for forecast:

•The presidential election has a positive impact in reducing the frequency of attacks compared to the year without presidential election. This makes sense as the security measures will be greatly increased for conduction smooth flow of elections.



	01Jan2013	01Jan2014	01Jan2015	01Jan2016	01Jan2017	01Jan2018	01Jan2019	01Jan2020
* US_POP_MILL	316.5	318.86	321.22009029	323.58015057	325.94021085	328.30027112	330.6603314	333.02039167
* US_PRESIDENTIAL_ELECTIONS	0	0	0	1	0	0	0	1
* US_DEF_BUDGET_MILL	769.93	761.5	600	600	600	731.03987895	726.23442008	722.14868821
* US_GDP_PERCENT_GR	2.2	3.66	2.0965668093	2.0597967778	2.0230267464	1.986256715	1.9494866835	1.9127166521
Model Forecast	980123652	14.562493545	14.804013501	17.737307355	15.506210877	16.676862383	15.829475105	16.211564524
Scenario Forecast	.	.	25.553587349	22.708096128	25.340882864	17.866347205	17.018972076	12.116769679

Figure 15: Scenario 2

•Scenario with change in budget from the above scenario, reduced the defense budget to 600 mill for 2015-17 to observe its effect on the frequency of attacks.

•Decrease in the defense budget resulted in an increase in number of attacks for the scenario years of 2015-17, even for an presidential election year, the number of attacks has increased greatly because of reduce in defense budget.

CONCLUSIONS

•Our Forecasting model predicts that there may be at least 16 attacks every year for the next 4 years

•From our study of factors which contribute to the success of the terrorist attack, we discovered that attack type, weapons used in the attack, place of attack and type of target play pivotal roles in determining the success of a terrorist attack

•Results reveal that the government may be successful in averting assassination attempts but may fail to prevent armed assaults and facilities/infrastructure attacks.

•So additional security may need to be provided to important facilities to prevent further attacks from being successful.

•Also results further reveal that it is possible to reduce the forecasted number of attacks by raising defense spending and by putting an end to the raging war in middle east.

REFERENCES

•http://www.usgovernmentspending.com/us_gdp_history

• http://www.multpl.com/united-states-population/table

• http://www.usfederalbudget.us/budget_pie_gs.php

• http://www.start.umd.edu/gtd/contact/

ACKNOWLEDGEMENT

We thank Dr. Goutam Chakraborty, Professor, Department of Marketing and Founder of SAS and OSU Data Mining Certificate Program – Oklahoma State University for his support throughout this research.

SAS® GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

LAS VEGAS | APRIL 18-21

#SASGF