Early Warning System for Quality Control

STUDENTS / UNIVERSITIES

Ibtissam Liedri / ENSIIE

Çiğdem Kenar / METU

SUPERVISOR(S)

Assoc. Prof. Kemal KILIÇ

ABSTRACT

Items forecasting is a sensitive activity that affects many of the company's operations. In our case study, each appliance that the manufacturer produces is an assemble where parts are supplied from various sources. Some of these suppliers are changed each month. In fact, forecasting how many items will demand service due to failure is closely related to modeling a product behavior and the relative complexity of the markets studied. These failures are influenced by many factors such as the month of the item production, the total time of usage, the season of the year as well as the region where the product has been used. It is therefore important for businesses to anticipate these types of failures for a more strategic internal corporate planning purposes and by that deliver the right product at the right time.

OBJECTIVES

- Predict how many items will fail in the upcoming months. The prediction will be done by setting up a model which considers the influence of the most important factors.
- Develop a system that will red flag the quality department whenever a problem occurs in the recent production.

DATASETS

Production Period	Production Amount	1. Ay	2. Ay	3. Ay	4. Ay	5. Ay	6. Ay	7. Ay	8. Ay
2012 Jan	84454	1,686	7,599	15,318	23,744	34,394	44,731	53,125	59,127
2012 Feb	100198	1,801	10,801	20,799	32,248	45,064	55,741	63,480	68,092
2012 Mar	102168	2,938	12,790	25,663	40,380	55,120	65,565	71,890	76,827
2012 Apr	76682	2,846	12,214	24,401	37,625	46,355	51,360	54,724	56,162
2012 May	67959	3,683	17,840	30,434	38,829	44,033	48,036	49,588	50,368
2012 Jun	61578	6,441	22,179	31,152	35,975	39,325	41,202	42,162	43,013
2012 Jul	80469	9,212	32,211	44,120	50,627	53,084	54,602	55,711	56,577
2012 Aug	91922	14,551	37,793	50,798	55,959	58,754	60,661	62,287	64,059
2012 Sep	90024	10,090	40,687	48,771	52,631	55,205	57,346	59,719	62,392
2012 Oct	97300	18,936	35,666	43,432	48,431	52,127	56,272	60,730	65,670
2012 Nov	70240	5,947	17,160	22,997	27,313	31,691	36,404	41,452	45,175
2012 Dec	53543	4,272	9,215	12,942	17,206	21,491	26,097	29,975	34,053
2013 Jan	80282	6,327	13,673	22,012	29,559	37,449	44,064	51,650	56,862
2013 Feb	67378	1,672	9,936	18,518	26,591	32,395	38,622	42,735	46,464
2013 Mar	126469	3,070	16,123	31,988	46,226	63,024	75,180	84,670	93,213
2013 Apr	125281	3,156	20,223	37,110	53,289	65,583	77,074	86,187	88,848
2013 May	122771	4,804	22,353	42,110	59,863	74,318	83,761	86,662	88,325
2013 Jun	88011	8,510	28,023	41,860	51,543	57,887	59,718	60,824	61,736
2013 Jul	113469	17,941	41,722	61,148	73,261	77,334	79,880	81,737	83,264
2013 Aug	67626	5,636	25,333	35,339	38,574	40,430	42,024	43,440	44,805
2013 Sep	113777	16,686	44,741	54,605	59,558	63,507	66,505	69,773	73,479

Two cumulative datasets:

- The failures data and which represent the number of items that failed in the months following their production
- The installations data in the months following their production.

METHODOLOGY

- ✓ **Study on the datasets:** Observing different factors such as seasonality, the usage period of an item etc.
- ✓ **Data cleaning**
- ✓ **Data transformation:** Different approaches are tried to merge the two datasets.
- ✓ **First model building**
- ✓ **Model improvement:** Feature subset selection and different encoding combinations are applied.
- ✓ **Performance measurement:** MAPE (Mean Absolute Percent Error), RMSPE (Root Mean Square Percentage Error), error in terms of items
- ✓ **Model selection**

RESULTS & CONCLUSION

		Without Feature Subset Selection		With Feature Subset Selection	
		RMSPE	MAPE	RMSPE	MAPE
Model 1	Random Forest	22.46%	12.9374	28.99%	13.0565
	Support Vector Regression (SVR)	51.84%	49.6901	51.84%	49.6901
Model 2	Random Forest	37.65%	11.4537	35.59%	11.7047
	Support Vector Regression (SVR)	51.84%	49.6901	51.84%	49.6901

Here, **Model 1** corresponds to the model with one-hot encoded months and years, whereas **Model 2** corresponds to one-hot encoded seasons.

Random forest applied to the second model with feature subset selection gives the best performance.

SVR performs the same for both models, with or without feature subset selection.

REFERENCES

- ANH QUOC DOAN, A. (n.d.). *Intro To Random Forest*. Lecture. Retrieved from https://web.csulb.edu/~tebert/teaching/lectures/551/random_forest.pdf
- Paisitkriangkrai, P. (n.d.). *Linear Regression and Support Vector Regression*. Lecture. Retrieved from https://cs.adelaide.edu.au/~chhshen/teaching/ML_SVR.pdf
- Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research,3*, 1157-1182.