OCS Accident Forecasting Using Data Mining Revisited

July 2017

The *Deepwater Horizon* accident increased the interest in accident prevention on the Outer Continental Shelf (OCS). One idea suggested by the Department of the Interior's Ocean Energy Safety Advisory Committee for the Bureau of Safety and Environmental Enforcement (BSEE) was an early warning system for accidents. Such system would fit within the BSEE goal of risked based decision making. In April 2014, a research project presented two such early warning systems (http://tedspublicpolicy.weebly.com/research-projects.html). Now that it is three years later, it is time to see how well these two early warning systems actually worked.

Most of the forecasting data used for this analysis is available from the BSEE data website (https://www.data.bsee.gov). The exception is the Incident of Non-Compliance (INC) which was obtain via Freedom of Information Act (FOIA) requests. That data is posted at this location: (http://tedspublicpolicy.weebly.com/research-projects.html). Both systems use this year's data to project the potential of accident in the following year. For this analysis data from 2013, 2014, and 2015 was used to project the accidents in 2014, 2015, and 2016.

The first system is a categorical algorithm. Six yes or no attributes of a specific lease year are examined. The number of yes is an indicator of the potential of an accident in the following year. The six questions are:

- Was there an INC during the year?
- Was there an INC of type E100, G110, P103 or P451 during the year?
- Was there an accident during the year?
- Was there production from 4 or more completions during the year?
- Was there 3 or more wells spud during the year?
- Was the lease located in any of these Area Codes: WR, KC, PL, GC, MC, SP, SS, MP, EW?

Here are the results:

Yes Count	No Accidents	Accidents	Percentage
0	954	3	0.31
1	1374	49	3.44
2	721	31	4.12
3	326	25	7.12
4	133	13	8.90
5	24	3	11.11
6	1	0	0.00

In the data table the rows are for specific number of Yes count. In the first row (Yes Count = 0) there were 957 (954+3) lease year combinations. Recall that the prediction data was from 2013, 2014, and 2015. In the following year there was no accident for that lease in 954 cases. There were 3 times an accident occurred. Those lease year combination with 0 Yes attributes had accidents about 0.31% of the time. While leases with 5 or more Yes attributes had accidents over 10% of the time. The categorical algorithm does distinguish between leases with high and low risk of accidents.

The second system is a logistic regression algorithm. The system computes the probability of an accident using 8 different lease year attributes. The prior paper describes the details of the math. The probabilities are rounded to the nearest whole percentage for aggregation. The results are:

Prob Interval	No Accident	Accident	Percentage
1	933	3	0.32
2	1245	36	2.81
3	397	20	4.80
4	194	13	6.28
5	284	19	6.27
6	66	4	5.71
7	109	5	4.39
8	35	2	5.41
9	101	7	6.48
10	28	4	12.50
11	23	1	4.17
12	46	3	6.12
13	8	2	20.00
14	24	1	4.00
15	6	1	14.29
16	2	0	0.00
17	5	1	16.67
18	3	0	0.00
19	3	0	0.00
20	3	0	0.00
21	5	1	16.67
22	1	0	0.00
23	4	0	0.00
24	1	0	0.00
25	0	1	100.00

The logistic regression estimates a probability that an accident will occur next year for each lease year combination. That probability is converted to a percentage and rounded to a whole number. That whole number is the first column labeled as Prob Interval. The next two columns display the count of No Accident and Accident observations in the following year.

The logistic regression algorithm does identify lease risks of accidents on a graduated scale. The following graph shows for projected probabilities up to 10 percent of an accident. The logistic regression algorithm is a good predictor. Above 10 percent that limited observations have a larger variation between the projection and the observed frequencies.

With three years of actual experience, both systems of accident forecasting do work effectively. The proof of concept has been confirmed. There is reason to believe that these algorithms can be improved:

- The data used in this approach was either published on the website or obtained under the FOIA process. BSEE has a much larger set of data from which insights can be made.

- No specific subject matter expertise was employed in the development. Insights from the BSEE professional staff should be of value.

- No updating of the forecasting system was made to reflect more recent data.

Data Mining approach to Accident Forecasting does work and there are reasons to believe even better systems can be developed.

Ted D. Tupper

PO Box 445
Summerdale, PA 17093