

CARLETON UNIVERSITY

SCHOOL OF
MATHEMATICS AND STATISTICS

HONOURS PROJECT



TITLE: A Mortality Experience Study using Machine Learning Methods

- Based on Canadian data 1991-2018

AUTHOR: Cameron McMullen

SUPERVISOR: Shirley Mills

DATE: April 30th, 2021

Table of Contents

Abstract	3
Chapter 1 – Introduction	4
Chapter 2 - Mortality Experience for Life Insurers	4
Chapter 3 - Machine Learning Algorithms for Regression.....	7
Chapter 4 - Introducing the Data	12
Chapter 5 - Data Visualizations	15
Chapter 6 - Modelling	21
Chapter 7 - Constructing Mortality Tables	41
References	43

Abstract

In this project, I do an experience study on the mortality of Canadians from 1991-2018. Leveraging the machine learning algorithms of linear regression, gradient boosting regression and support vector regression, I forecast the one-year death rates in the year 2018 by using the mortality experienced from 1991-2017.

1. Introduction

Machine Learning has been a rapidly growing field in today's world. Many of the latest and greatest innovations in the technology industry have some artificial intelligence component and every business is trying to figure out a way to leverage the power of data science to advance their products.

One industry where the application of data science is being integrated is life insurance. When life insurers perform the underwriting process to evaluate a person's level of risk, they collect a lot of personal data. The data is used in actuarial models to calculate the premiums and the reserves. Another purpose for collecting personal data is for predicting a person's mortality rate. Mortality rates measure the rate of death in a population over time. Mortality rates are a critical component that is inputted into the pricing and valuation models.

Deriving an accurate mortality curve is complicated and requires analysis of the mortality experience, the rate of deaths that have been experienced recently in time. In practice, there is no distribution that completely defines the mortality rate of a population. Everybody is unique. However, there are similarities that are shared between groups of people that can help define a distribution of best fit. Machine learning methods can be used to uncover the distribution of best fit.

The second problem with mortality experience is that the best fit distribution is constantly changing over time. For example, the mortality of the global population has heavily declined since the start of the 20th century because of technological advancements. To take this into consideration, actuaries typically do mortality improvement studies. The purpose of this study is to update the old mortality rates in the valuation models to a "better" set of mortality rates.

In general, Insurance companies store massive amounts of data. With this data, they can use machine learning and data science to solve many problems. In this report, we investigate mortality rates for the Canadian population in 2018 by conducting a mortality experience study using three (3) machine learning methods.

2. Mortality Experience for Life Insurers

At the heart of any life insurer's business are life contingent risk models. A life insurer sells life insurance products. The cost of the life insurance products depends on the life expectancy of the insured. To buy an insurance policy, one typically agrees to pay a sequence of cash flows (CF) over a period of time in exchange to have their beneficiary receive a lump sum of money on the occurrence of their death. The cashflow payments made from the policyholder are called premiums and the lump sum paid out by the insurer is typically called the death benefit. There are many different payment structures and products for different coverage periods. Figure 1 is a basic diagram of the cash flows from the perspective of the insurer.

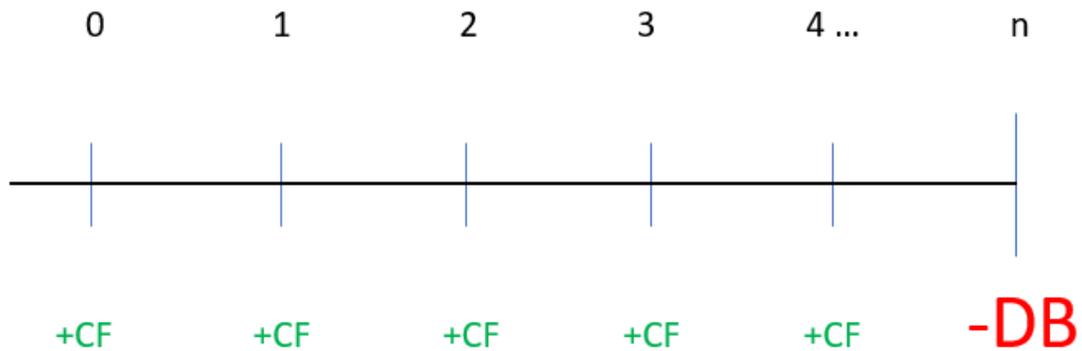


Figure 1. Example of cash flows (CF) from perspective of insurer. +CF is a premium received by the insurer, -DB is the death benefit paid out by the insurer.

It is critical that the actuarial models for mortality are accurate as the profitability of the business strongly depends on them. One of the simplest ways to verify the validity of the models is to compare the actual claim amounts to the expected claim amounts. The expected claim amounts are produced using actuarial assumptions and known data inputted into the model. Mortality is one of the key assumptions needed to produce the expected claim amount for a given policy.

The age at death of a person is a random variable whose cumulative distribution function (cdf) must satisfy three (3) properties:

Let X be a random variable for the age at death.

1. $F(0) = 0$

2. $F(X)$ is a non decreasing function

3. $\lim_{x \rightarrow \infty} F(x) = 1$

Every person has a unique mortality distribution. To simplify things, actuaries classify people into different risk classes. For example, smokers and non-smokers would have significantly different mortality distributions, therefore it is appropriate to have different mortality rates for these groups of people. Other variables that are commonly used for distinguishing risk classes are:

- Sex
- Face amount of the Policy
- Age
- Country where they live

The objective is to then create a probability mass function (pmf) for the age at death given these input variables.

$$P_X(x|v_1, v_2, \dots, v_n), \quad x = 0, 1, 2, \dots$$

where x is the age at death and v_1, \dots, v_n are the n input variables.

In practice, the distribution of the age at death is discrete. Actuaries record the distribution in the form of a table. Below is an example of a mortality table, where l_x represents the number of survivors at age x . Table 1 directly gives the survival function $P(X \geq x) = \frac{l_x}{l_0}$ of the life random variable.

Table 2000CM

Age x	l_x	Age x	l_x	Age x	l_x
0	100000	37	96921	74	66882
1	99305	38	96767	75	64561
2	99255	39	96600	76	62091
3	99222	40	96419	77	59476
4	99197	41	96223	78	56721
5	99176	42	96010	79	53833
6	99158	43	95782	80	50819
7	99140	44	95535	81	47694
8	99124	45	95268	82	44475
9	99110	46	94981	83	41181
10	99097	47	94670	84	37837
11	99085	48	94335	85	34471
12	99073	49	93975	86	31114
13	99057	50	93591	87	27799
14	99033	51	93180	88	24564
15	98998	52	92741	89	21443
16	98950	53	92270	90	18472
17	98891	54	91762	91	15685
18	98822	55	91211	92	13111
19	98745	56	90607	93	10773
20	98664	57	89947	94	8690
21	98577	58	89225	95	6871
22	98485	59	88441	96	5315
23	98390	60	87595	97	4016
24	98295	61	86681	98	2959
25	98202	62	85691	99	2122
26	98111	63	84620	100	1477
27	98022	64	83465	101	997
28	97934	65	82224	102	650
29	97844	66	80916	103	410
30	97750	67	79530	104	248
31	97652	68	78054	105	144
32	97549	69	76478	106	81
33	97441	70	74794	107	43
34	97324	71	73001	108	22
35	97199	72	71092	109	11
36	97065	73	69056	110	0

Table 1. Mortality table 2000CM taken from [Link](#)

3. Machine Learning Algorithms for Regression

This chapter will explore three (3) machine learning algorithms that will be implemented in the analysis section to follow. The algorithms that will be introduced are:

1. Linear Model
2. Gradient Boosting Decision Tree
3. SVM (Support Vector Model Regression)

3.1 Linear Model

Linear regression is a statistical technique that is commonly used to construct machine learning models. The estimated model is of the form:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p$$

- \hat{y} is the response variable
- x_i 's are the features, $i=1, \dots, p$
- p is the number of features
- $\hat{\beta}_i$'s are the estimators for the coefficients

One assumption of the model is that the features need to be numerical variables. To handle categorical features with n layers, that is, a feature that is one of n possibilities, dummy variables are created. Dummy variables are indicator variables, they indicate whether an element is in a set or not. For a categorical feature with n layers, a transformation will be applied so that there are $n-1$ indicator variables each with two layers. This will create a regression model in the form:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m + \hat{\theta}_1 q_1 + \dots + \hat{\theta}_{n-1} q_{n-1}$$

- \hat{y} is the response variable
- x_i 's are the numerical features, $i=1, \dots, m$
- m is the number of numerical features
- $\hat{\beta}_i$'s are the estimators for coefficients, $i=1, \dots, m$
- q_j 's are indicator variables, $j=1, \dots, n-1$
- θ_j 's are estimators for the coefficients of the indicator variables
- $n-1$ is the number of indicator variables needed to represent one categorical variable having categories

3.1.1 Ordinary Least Squares (OLS)

Ordinary least squares is a method used to find the best estimators for the coefficient vector. The best coefficients for the model are the ones that minimize $\|\hat{y}_i - y_i\|^2$

In matrix notation:

$$\|\hat{y}_i - y_i\|^2 = \sum_{i=1}^n (\mathbf{x}_i^T \boldsymbol{\beta} - y_i)^2 = (\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta})$$

where

$$X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \cdots & X_{pn} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \theta_{n-1} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

If the columns of matrix X are linearly independent, there is a unique solution to the minimization problem, given by the explicit formula:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{n+1}}{\operatorname{argmin}} (\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta}) = (X^T X)^{-1} X^T \mathbf{y}$$

Therefore, given the matrix X of the features and the respective response variable values \mathbf{y} , a linear regression model can be created by the above formula.

3.2 Gradient Boosting Decision Tree

3.2.1. Decision Trees

A decision tree is a visual tool to break down a dataset into smaller groups. A decision tree consists of leaves and decision nodes. Each leaf is made up of points from the data with similar parameters. The decisions nodes show how the data is split apart. [Figure 2](#) below is an example of a decision tree:

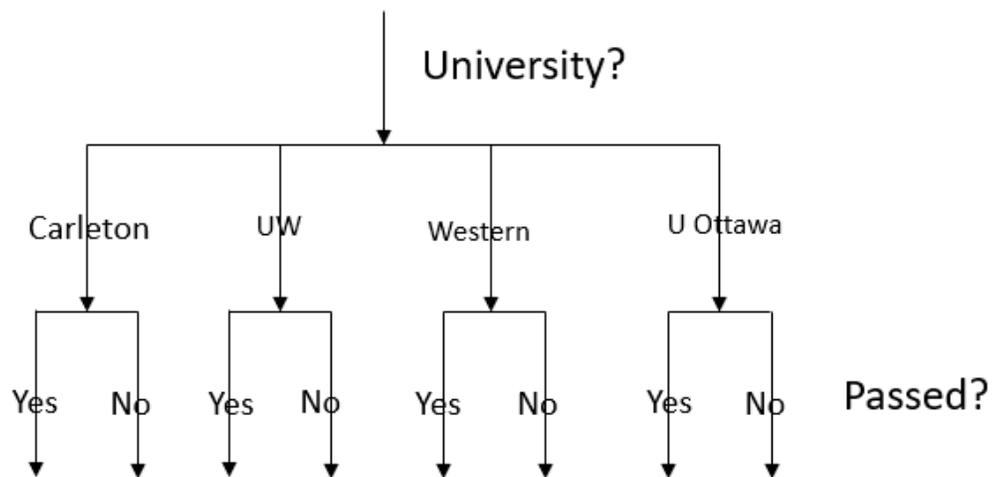


Figure 2. Example of a Decision Tree

There are four (4) leaves at the first layer of the tree, one for each university. The decision node is the university the person attends. At the second layer there are two (2) leaves for each branch. The decision node is whether they passed or not. From the tree, the data is partitioned into eight (8) different groups since the tree has eight (8) leaves.

There are certain machine learning algorithms that are implemented on decision trees. The optimal number of decision nodes and leaves depends on the dataset, the machine learning algorithm and the criterion used to measure purity or error at a node.

3.2.2. Random forests (RF)

Random forests is a machine learning technique that uses multiple decision trees. Decision trees are created from random samples of the data. A machine learning algorithm is then implemented on each tree, to try and predict the result for a case. Finally, weightings are given to each tree based on the accuracy of its model prediction. A final formula that maps the feature set of variables to the response variable using the weightings is then produced.

3.2.3. The following is a general procedure for the Gradient Boost algorithm. X represents the predictors and y represents the response variable.

1. Fit a decision tree f_1 so that $f_1(x) = y$

2. Fit a decision tree r_1 to the residuals of the model in (1), so that $r_1(x) = y - f_1(x)$, the residuals of the model f_1
3. Refine the model to be $f_2(x) = f_1(x) + r_1(x)$
4. Fit decision tree r_2 to the residuals of the model in (3) so that $r_2(x) = y - f_2(x)$, the residuals of the model f_2
5. Refine the model to be $f_3(x) = f_2(x) + r_2(x)$
6. Repeat this process until a desired stopping condition(s) is met

The following is an in-depth example of using the gradient boost algorithm.

1. Start with a root node that has a single value. It represents the initial estimate for the response, y , of any given sample. The first estimate is the average of all the responses in the data.

$$\text{i.e. calculate } \widehat{y}_1 = \frac{\sum_{i=1}^n y_i}{n}$$

2. Determine the residuals from all observations and the estimator \widehat{y}_1

$$\text{i.e. calculate } r_i = y_i - \widehat{y}_1 \quad \forall i = 1, \dots, n$$

3. Partition the sets of predictor variables into groups. For example, assume each response y has three predictor sets A, B, C, such that $\widehat{y} = f(a, b, c)$ where $a \in A, b \in B, c \in C$. We partition each predictor set A, B, C into a group of k mutually exclusive sets such that for each set:

- a. $\bigcup_{i=1}^k A_i = A, \bigcup_{i=1}^k B_i = B, \bigcup_{i=1}^k C_i = C$
- b. $\bigcap_{i=1}^k A_i = \emptyset, \bigcap_{i=1}^k B_i = \emptyset, \bigcap_{i=1}^k C_i = \emptyset$

This will create j sets of residuals. Each set of residuals P_j are derived from the observations that share the same predictor variable partition sets, i.e.,

$$P_j = \{r_{ij} | a \in A_l, b \in B_m, c \in C_n\}, j=1, \dots, k^3$$

where a, b, c are the predictor variables for response observation y_{ij} in prediction partition j and A_l, B_m and C_n are the individual partition sets corresponding to partition j .

4. Take the average of the residuals in each set P_j , i.e.,

$$R_j = \bar{P}_j = \sum_{i=1}^n \frac{r_{ij}}{n}$$

5. An estimator for y can then be created from its set of predictor variables $\{a, b, c, \dots\}$ by the formula:

$$\widehat{y}_2 = \widehat{y}_1 + R_j$$

Figure 3 below presents a visualization of the partitioning with $k = 2$.

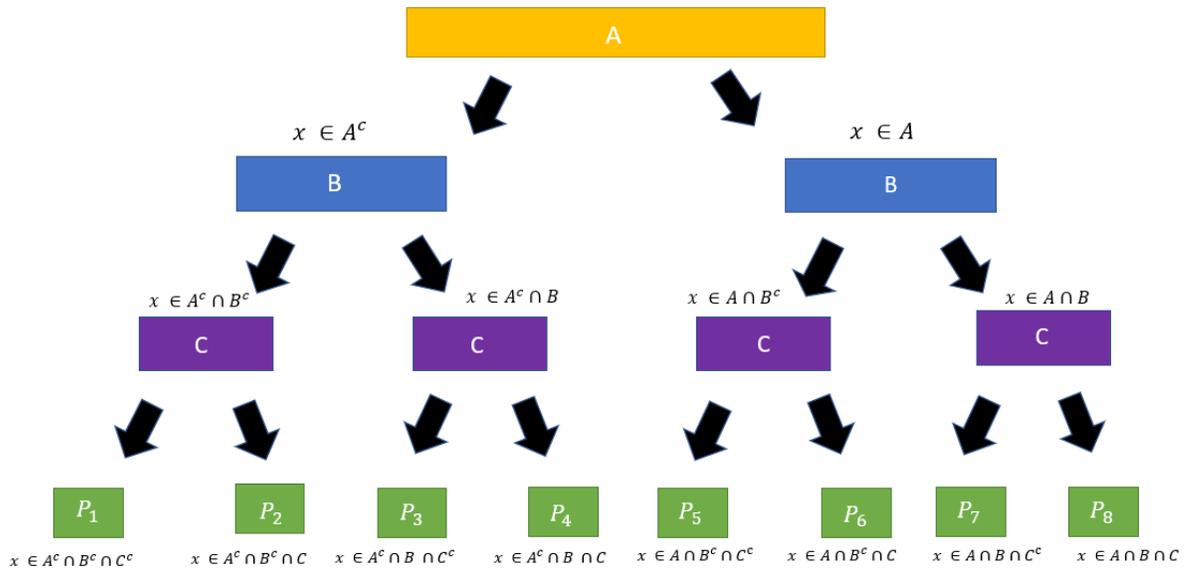


Figure 3. Visual representation of partitioning

3.3 Support Vector Regression Model (SVM)

Just like other regression techniques, SVM produces a function $f(\mathbf{x}) = y$, where \mathbf{x} is the vector of predictor features and y is the response. This is a convex optimization problem that tries to minimize:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

under the constraints:

$$\begin{cases} |y_i - \mathbf{x}_i^T \mathbf{w}| \leq \varepsilon + \xi_i \\ \xi_i \text{ and } \xi_i^* \geq 0, i = 1, \dots, n \end{cases}$$

where \mathbf{w} are the weights, ε is a threshold that is set, ξ_i and ξ_i^* , are distances to outliers beyond the threshold ε and C is a constant.

A visual representation for the linear case is given in Figure 4 below:

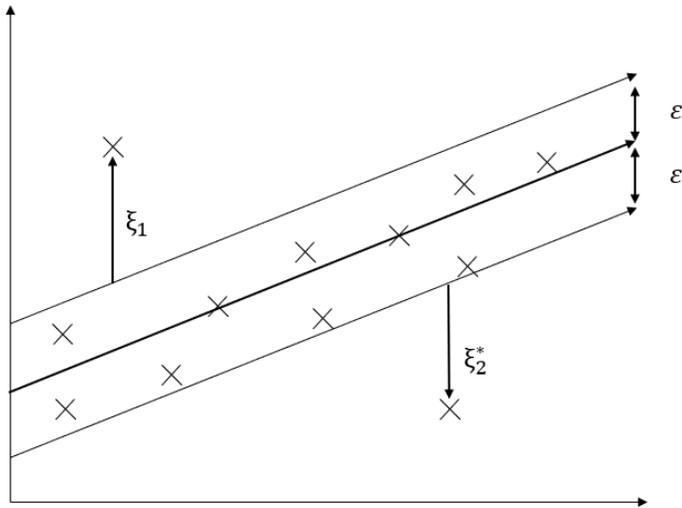


Figure 4. Diagram of a linear case for Support Vector Regression

This minimization problem can also be extended to other continuous functions besides a linear function.

Then the optimization problem is minimizing:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

Under the constraints:

$$\begin{cases} |y_i - (\phi(\mathbf{x}_i))^T \mathbf{w}| \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, n \end{cases}$$

where ϕ is a transformation applied to the predictor variable space.

4. Introducing The Data

The data (see Table 2 below) was constructed by combining data from [GoC](#) (Government of Canada) and [UW](#) (University of Waterloo). It shows the number of deaths in Canada from 1991-2018.

In total there are eight (8) fields:

1. Year – Year of the deaths
2. GEO – Province/Territory in Canada where people died
3. Age – The age at death. This is split into 6 buckets:
 - “<15”

- "15-44"
 - "45-64"
 - "65-74"
 - "75-84"
 - "85+"
4. Sex – Sex of the deceased.
 - Male
 - Female
 5. Marital Status – Marital status of the deceased, this is split into 6 buckets:
 - "Divorced"
 - "Married"
 - "Separated"
 - "Single"
 - "Widowed"
 - "Unknown"
 6. Deaths – Number of deaths
 7. Population – This is the population in Canada grouped by sex and age
 8. Percent of smokers – percent of people in Canada that are smokers grouped by sex only

Table 2: Example of the Data collected for analysis

Year	GEO	Age	Sex	Marital_Status	Deaths	Population	Smoking_percent
1991	Alberta	<15	Female	Divorced	0	2826334	30
1991	Alberta	<15	Female	Married	0	2826334	30
1991	Alberta	<15	Female	Separated	0	2826334	30
1991	Alberta	<15	Female	Single	200	2826334	30
1991	Alberta	<15	Female	Widowed	0	2826334	30
1991	Alberta	<15	Male	Divorced	0	2969226	32

From Table 2, the number of deaths are grouped into different risk classes. The attributes that make up each risk class are:

- GEO
- Age
- Sex
- Marital Status

Let $S_G, S_A, S_S,$ and S_M be the sample spaces for GEO, Age, Sex, and Marital Status respectively.

$|S_G| = 13, |S_A| = 6, |S_S| = 2, \text{ and } |S_M| = 6.$

Then, for a death, the total number of risk classes it can be placed in is:

$$|S_G| \times |S_A| \times |S_S| \times |S_M| = 13 \times 6 \times 2 \times 6 = 936$$

It is important to note that, from Table 2, the percentage of female smokers in 1991 in Alberta under age 15 was 30% and the percentage of similar male smokers in 1991 was 32%. In Table 2, smoking percentages do not change based on age, geographical region or marital status. It is also important to note that in Table 2, the female population in this partition does not change based on marital status and province.

Assumptions

Data for the “Smoking_percent” column initially had missing values for both males and females in the years 1992, 1993, 1997, 1998, 2012, 2013, and 2014. Linear interpolation was used to fill in the missing values. That is, if n is a year with known smoking percentage and $n + m$ is a year with known smoking percentage, where $m > 1$ and $n + i$ are years with unknown smoking percentages, $\forall i, 0 < i < m$ then:

$$s_{n+i} = s_n + \frac{i (s_{n+m} - s_n)}{m}$$

5. Data Visualizations

1. Number of Canadian deaths on average out of 100,000 people from the population of Canada

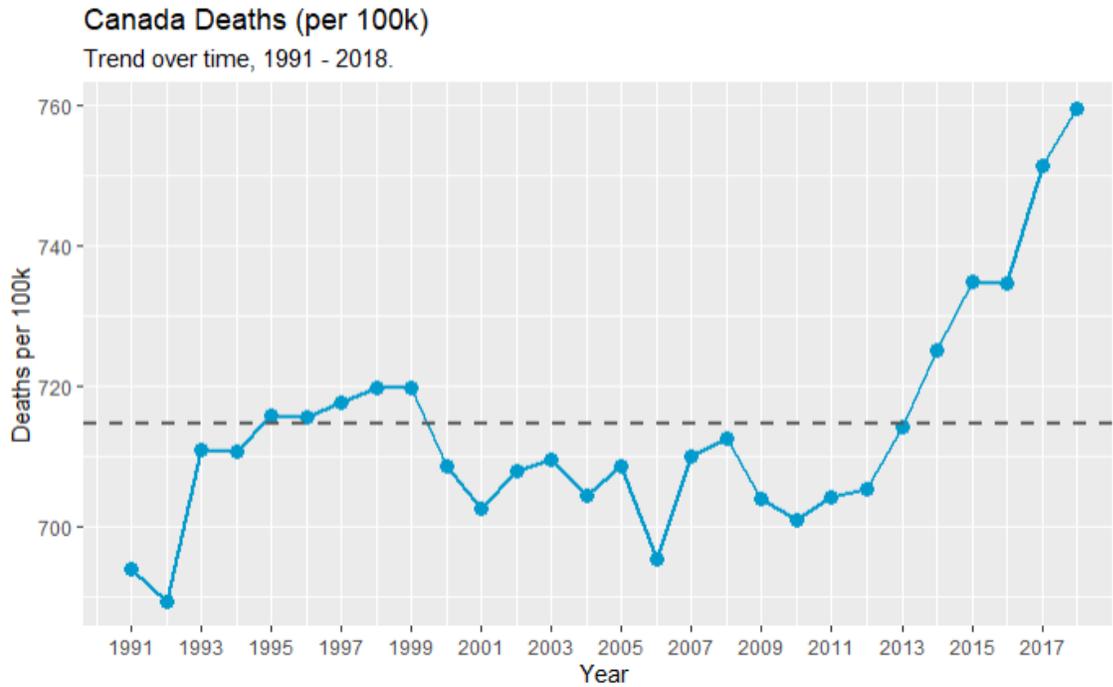


Figure 5. Death per 100k population in Canada, 1991-2018. Lines 202-219 of [Code](#).

Figure 5 shows that over time there has been an upward trend in death rates. The average number of deaths per 100,000 people of the Canadian population from 1991-2018 has been 715 deaths. (This rate has shown a significant increase between the years of 2013-2018.) The maximum death rate over this period is 760 deaths per 100,000 people and is seen in the year 2018. The minimum death rate over this period is 689 deaths per 100,000 people and is seen in the year 1992.

2. Average number of Canadian deaths per 100,000 people by sex

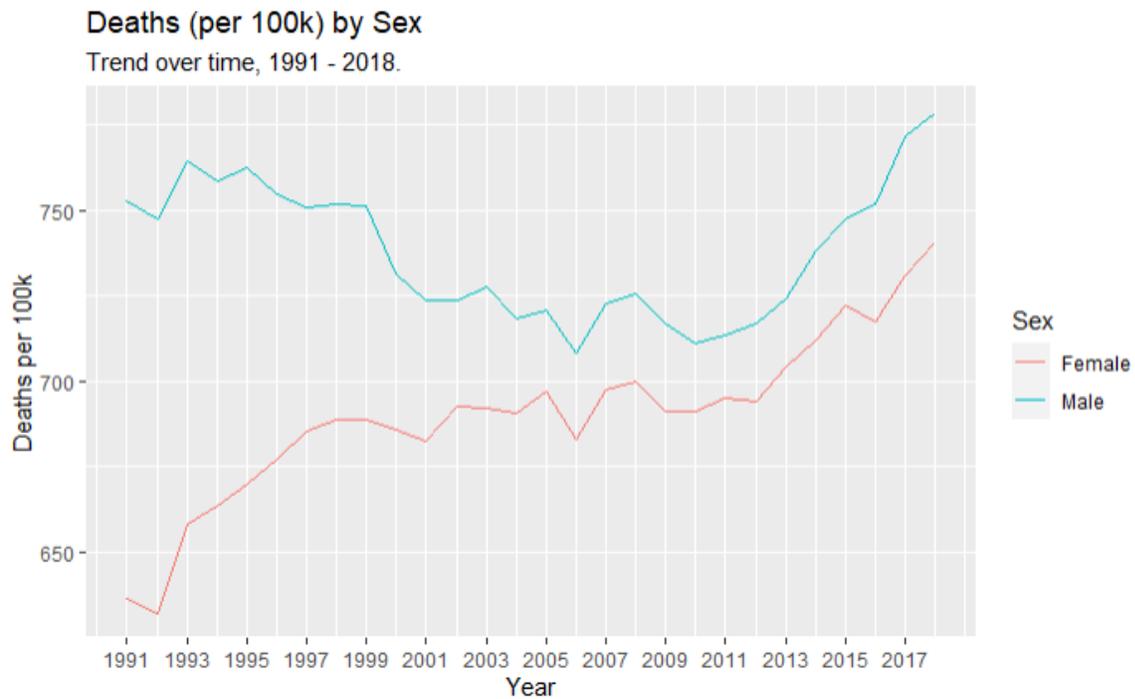


Figure 6. Deaths per 100k population In Canada 1991-2018, by sex. Lines 228-242 of [Code](#).

Figure 6 takes figure 5 and breaks deaths down by gender. For the period 1991-2018, in Canada male death rates have consistently been higher than female death rates over time. This supports the fact that Canadian females have a longer life expectancy.

3. Average number of Canadian Deaths by Age group per 100,000 People from the population of Canada

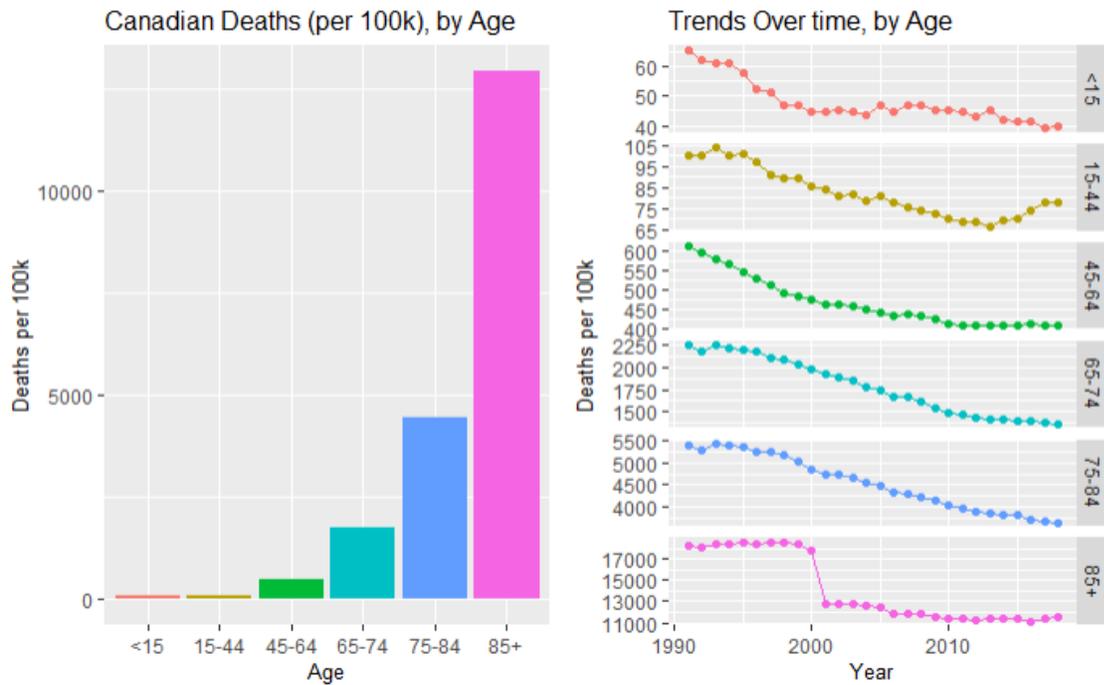


Figure 7. Average number of deaths in Canada 1991-2018 by age groups. Lines 255-293 of [Code](#).

Figure 7 above shows Canadian death rates have been decreasing over time for all age groups except for 15-44 year-olds. From 2013-2018 the death rate of 15-44 year-olds has been increasing. Since the size of the population that falls into the 15-44 age group is large, it is the driving force of why Canadian mortality rates have overall been increasing from 2013-2018. This is shown in Figure 5.

4. Total number of Canadian Deaths by Marital Status out of the total population of Canada

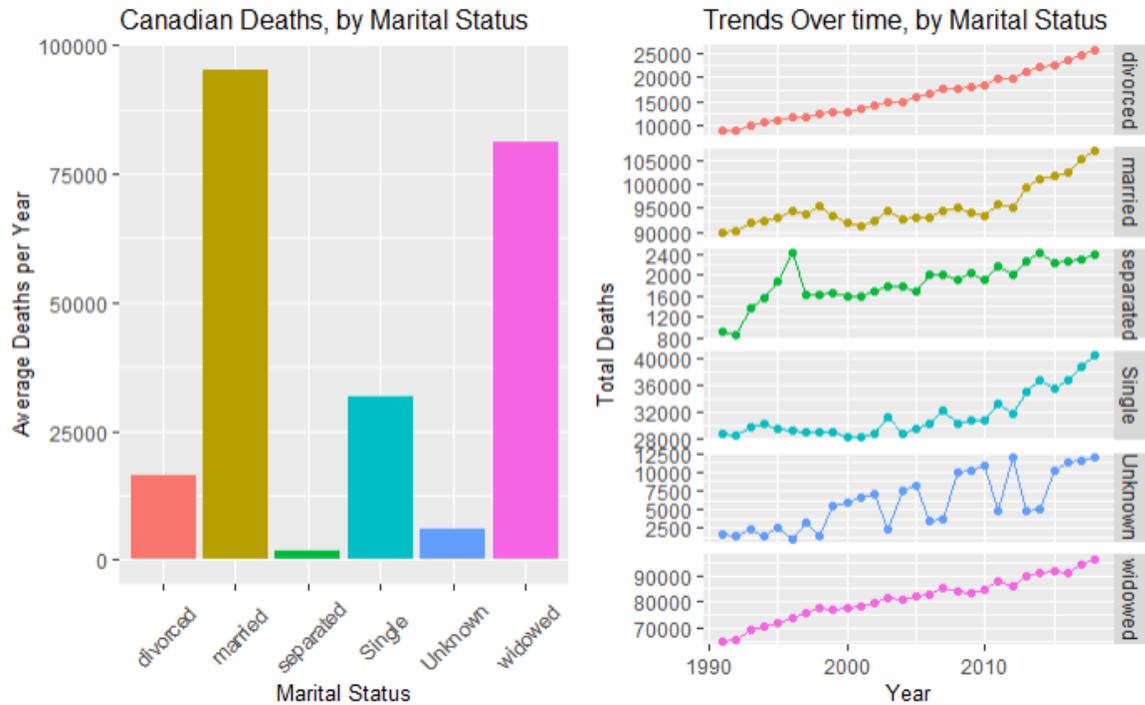


Figure 8. Deaths in Canada 1991-2019 by marital status. Lines 306-341 of [Code](#).

Of all the deaths that occurred in Canada from 1991-2018, [Figure 8](#) shows most were from people that were married. On average, 95,254 married people died every year. The total number of deaths across all six marital status groups has been increasing over time. The total number of deaths in the “Single” group has been exponentially increasing over the last few years. Other marital status groups have had total deaths linearly increasing over this period of time.

5. Population of Canada by Age over time

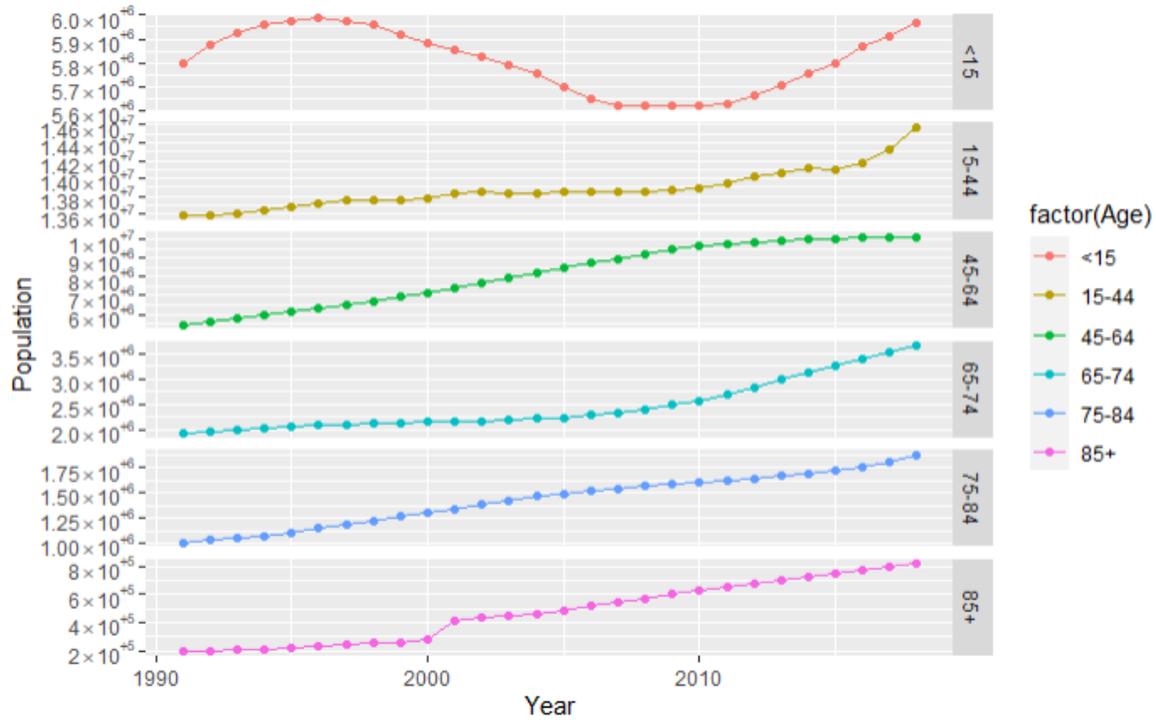


Figure 9. Population of Canada 1991-2018 by age groups. Lines 181-195 of [Code](#).

Figure 9 shows that the population for the older age groups have increased by a greater percentage than the younger age groups since 1991. The number of people under 15 years old has been fluctuating between 5.7 million and 6 million people. This is the only age group that does not have a clear upward trend since 1991; in fact its curve is decidedly nonlinear..

6. Total Canadian Deaths over time by Marital Status and Age

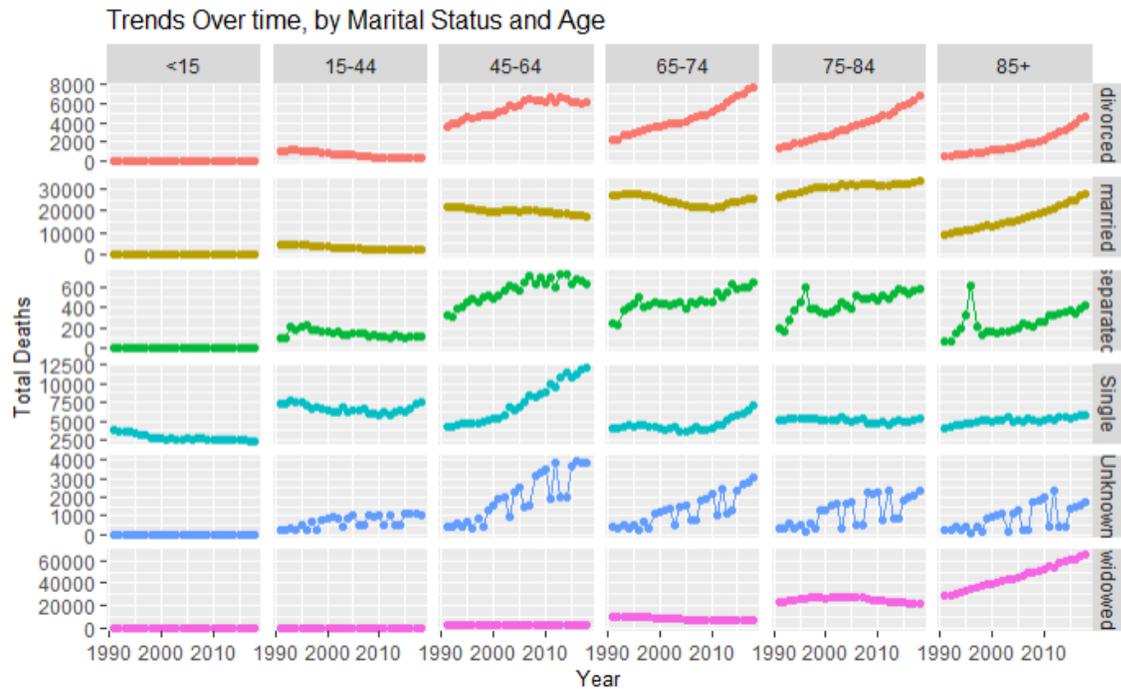


Figure 10. Canadian deaths 1991-2018 by marital status and age group. Lines 346-357 in [Code](#).

Figure 10 shows the total number of deaths over time by age group and marital status. One thing that stands out is the decreasing number of deaths for people 45-64 years old and married. This stands out because the population of the 45-64 year old age group has increased from 1991-2018 (see Figure 9). This could imply two things:

- a. That mortality rates for married people between the ages of 45 and 64 have decreased since 1991.
- b. That the number of married 45-64 year-olds has decreased since 1991.

The data does not contain the population of Canadians for each marital status so no conclusion can be made about either of these statements.

The second thing that stands out is the increase in number of deaths in singles aged 45-64. Some of this increase is explained by the doubling in population for this age group since 1991, but the deaths for this group has increased a staggering 2.8 times since 1991.

7. Percent of the Canadian population that smokes by gender

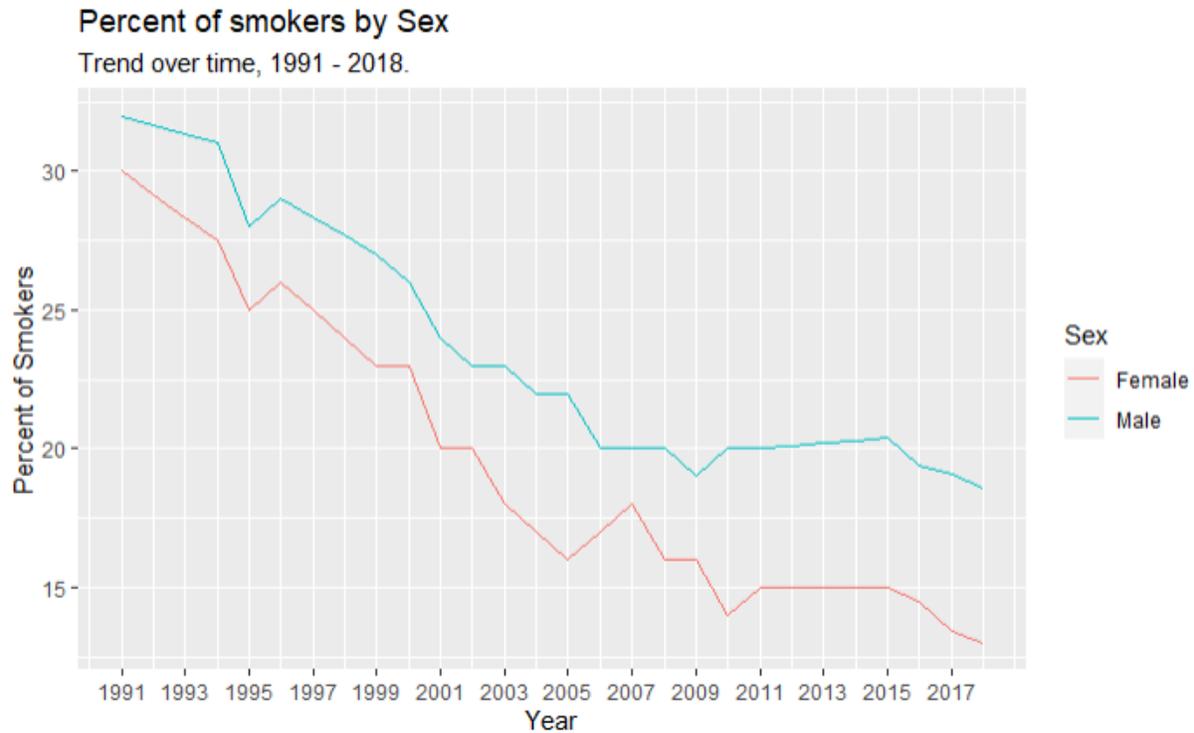


Figure 11. Percentage of smokers by sex in Canada, 1991-2018. Lines 365-377 in [Code](#).

From Figure 11 above we see the percentage of Canadians that smoke has been decreasing over this time period. Also, the percentage of males that smoke has been consistently higher than the percentage of females that smoke.

6. Modelling

This section will analyze the applications of the three (3) machine learning algorithms presented in [chapter 2](#). The algorithms are applied to the data (excluding 2018) with the goal of developing unique mortality tables for each risk class. The models are scored based on the accuracy of their prediction for the total number of deaths seen in Canada in 2018.

6.1. Problem

In order to construct a mortality table, we need an approximate distribution for a person's age at death. Specifically, we ask is: How many deaths can we expect in one (1) year for a certain risk class. In mathematical terms, this problem can be written in the form of a function.

$$f: \{A, G, S, M, S_p, Y, P\} \rightarrow \mathbb{Z}_{\geq 0} \quad (1)$$

where A, G, S, M, S_p, Y and P are the sample spaces for age, geography, sex, marital status, smoking percent, year and population, respectively.

This is a fundamental question in actuarial science because it is directly related to the one year survival probability, p_x . If the survival function is known, then the one year force of mortality can be estimated using the formula:

$$\mu_t = \frac{f_0(t)}{S_0(t)} = \frac{-S'_0(t)}{S_0(t)}$$

Where

μ_t = the force of mortality at time t

$S_0(t)$ = the survival function for a person age 0

$f_0(t)$ = the pdf of the age at death for a person age 0

The one year survival rate is also a critical component to the construction of a mortality table. A mortality table shows the exhaustion of a population. Below is the mortality table for an initial population of 1000 people. As age $(x) \rightarrow \infty, l_x \rightarrow 0$.

Table 3. Mortality Table example

Age (x)	l_x	P_x	d_x
0	1000	0.98	20
1	980	.	.
2	.	.	.
3	.	.	.
4	.	.	.
5	.	.	.
6	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

where

d_x = the number of deaths at age (x)

p_x = the probability of a person age (x) surviving the next year

l_x = the size of the population at age (x)

6.2. Linear model as a solution

The data introduced in [section 4](#) was used to construct a set of linear models that each predict the number of deaths in a risk class for the year 2018. The training data used to construct the models were all numbers from 1991-2017 that belonged to the same risk class.

Training data for risk class $C = \{x : A = a, G = g, S = s, M = m, 1991 \leq Y \leq 2017\}$. That is, the training data for a risk class consists of all cases with the same age, gender, sex, marital status and are from 1991-2017.

The training data for each risk class was used to construct a linear model for the number of deaths. The total number of deaths for the year 2018 was predicted using the invariance property as the sum of the predicted deaths for each risk class. That is:

$$\hat{y} = \sum_c \widehat{D}_c$$

where \hat{y} is the expected total number of Canadian deaths and \widehat{D}_c is the expected number of Canadian deaths in risk class c .

Recall policy holders in the same risk class share the same age group, sex, marital status and geo. After filtering the data for these criteria, a linear model can be created based on the predictors “Year”, “Population” and “Smoking_percent”. This creates the model:

$$\widehat{D}_c = \widehat{\beta}_{0c} + \widehat{\beta}_{1c}x_1 + \widehat{\beta}_{2c}x_2 + \widehat{\beta}_{3c}x_3 \quad (2)$$

where x_1, x_2 and x_3 are “year”, “population” and “Smoking_percent” respectively.

Table 4 below is an example of the filtered data, so that all deaths are from the same risk class. The risk class is Canadians that are:

1. In Ontario
2. 15-44 years old
3. Male
4. Single

Table 4. Example of deaths for risk class (single, male, 15-44, Ontario). Line 409 in [Code](#).

Year	GEO	Age	Sex	Marital_Status	Deaths	Population	Smoking_percent
1991	Ontario	15-44	Male	Single	1792	6871290	32
1992	Ontario	15-44	Male	Single	1835	6862997	31.7
1993	Ontario	15-44	Male	Single	1826	6867523	31.3
1994	Ontario	15-44	Male	Single	1774	6884694	31
1995	Ontario	15-44	Male	Single	1831	6905900	28

1996	Ontario	15-44	Male	Single	1669	6928486	29
------	---------	-------	------	--------	------	---------	----

Expanding off of expression (1), the model that predicts the number of deaths for a risk class can be decomposed into two functions:

$$f: \{A, G, S, M, S_p, Y, P\} \rightarrow \mathbb{Z}_{\geq 0} = g: \{A, G, S, M; S_p, Y, P\} \rightarrow h: \{S_p, Y, P\} \rightarrow \mathbb{Z}_{\geq 0} \quad (3)$$

where g is a function that takes age, gender, sex and marital status and maps it to another function h , h is the linear function shown in expression 3. It takes the population, year and smoking percent and maps it to the number of deaths in the risk class.

6.2.1. Results

Table 5. Predicted results in Canada for 2018 using linear model

Canada

Year	Actual Deaths	Expected Deaths	Percent Error
2018	283,706	275,406	3.0%

Overall, the results from this model were good. From Table 5 we see the total number of deaths in 2018 predicted by the model was 275,406, while the actual number of deaths was 283,706. This gives a percent error of 3%.

The square root of the average squared deviations was 58.2.

Table 6. Predicted results for Ontario for 2018 using linear model

Ontario

Year	Actual Deaths	Expected Deaths	Percent Error
2018	106,991	102,297	4.6%

The total number of deaths in Ontario in 2018 predicted by the model was 102,297, while the actual number of deaths was 106,991. This gives a percent error of 4.6%.

The square root of the average squared deviations was 156.8.

6.2.2. Visualization of Results

Below are some visualizations that display the accuracy of the linear models across all the risk class categories.

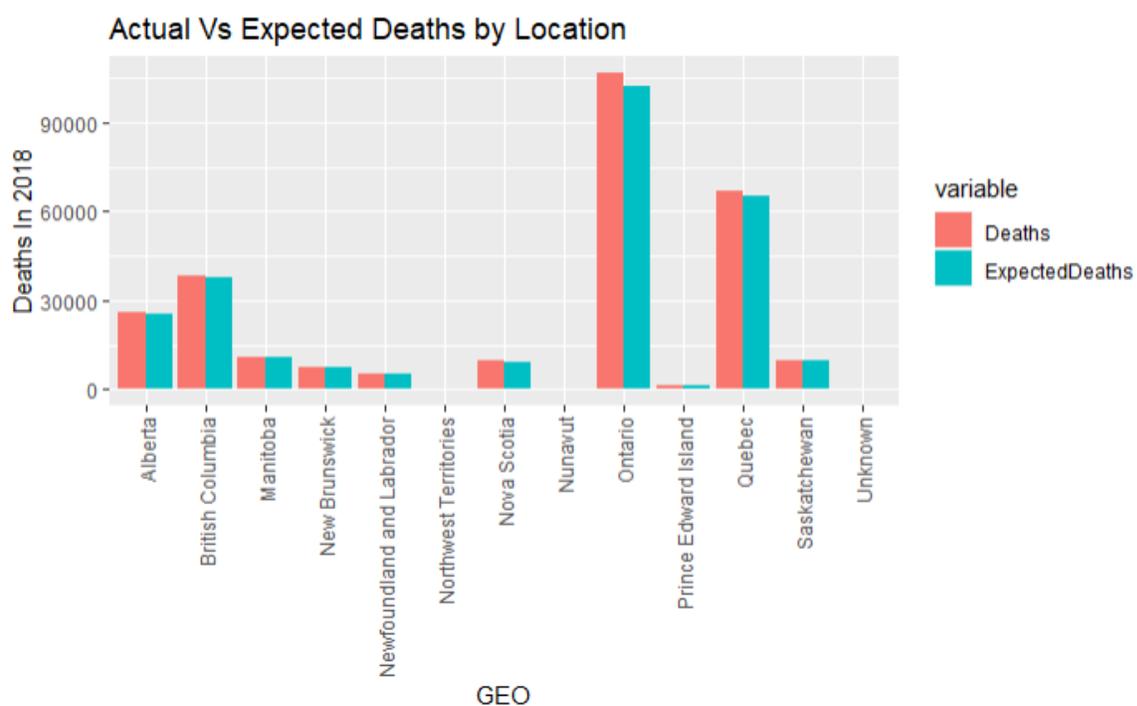


Figure 12. Actual vs. Expected Deaths in Canada in 2018 by Geographical Location, using linear models.
Lines 523-533 in [Code](#).

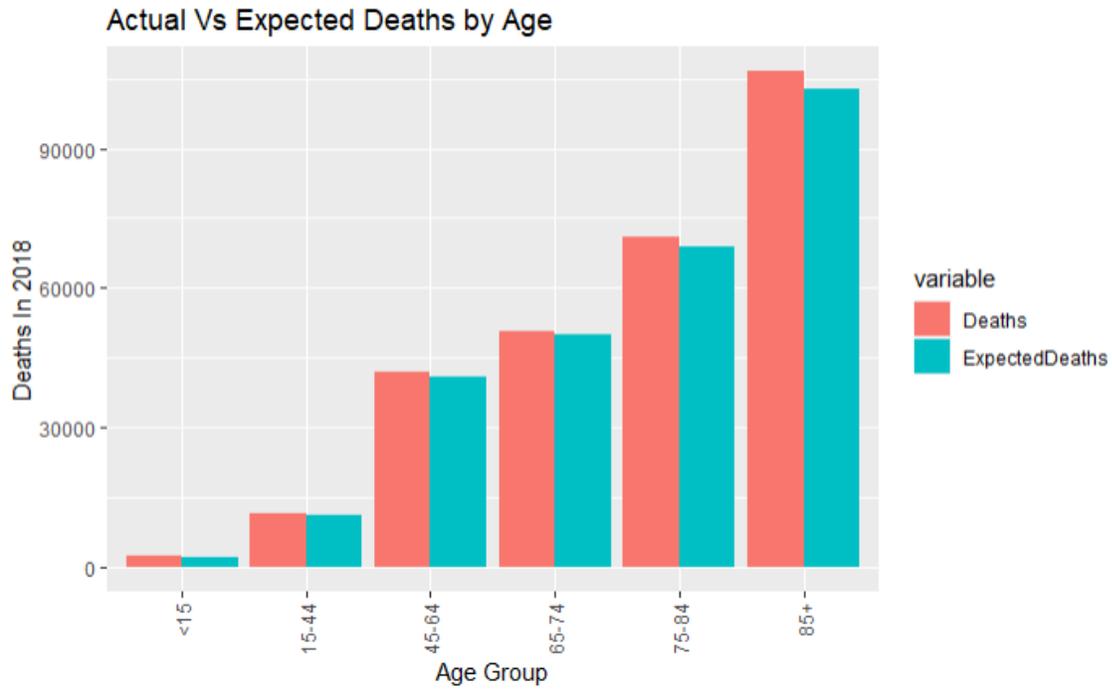


Figure. 13. Actual vs. Expected Deaths in Canada in 2018 by Age, using linear models. Lines 540-550 in [Code](#).

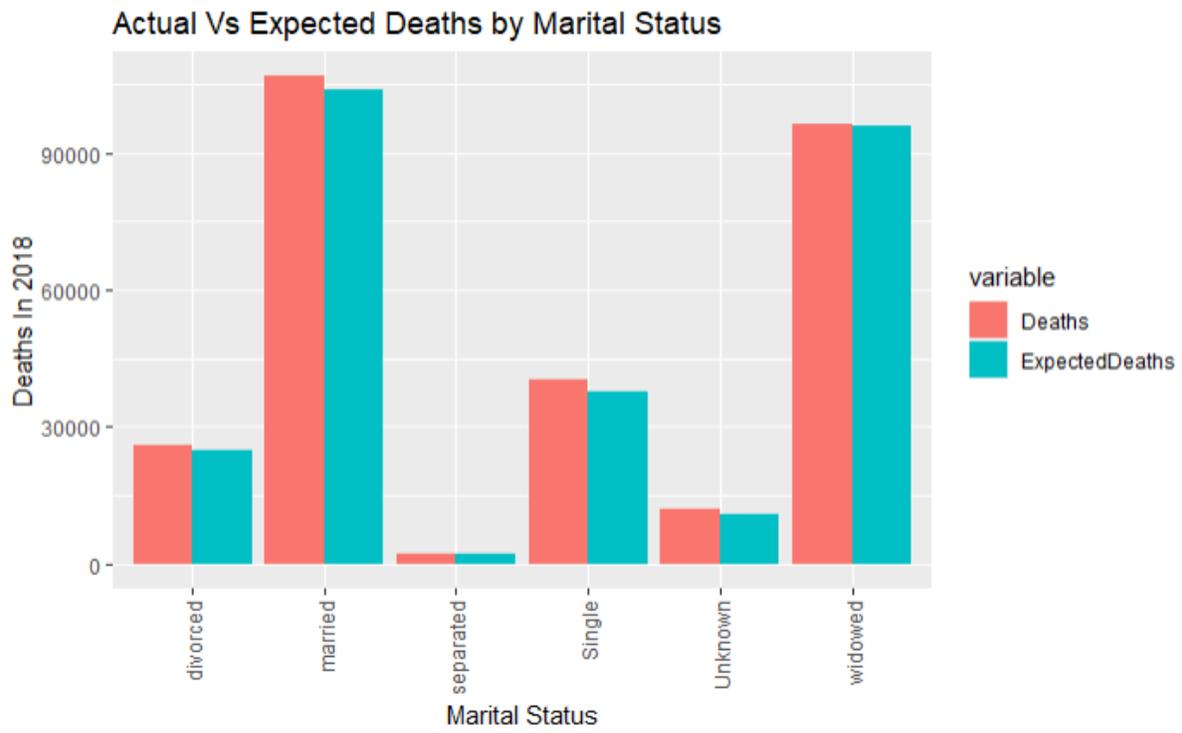


Figure. 14. Actual vs. Expected Deaths in Canada in 2018 by Marital Status, using linear models. Lines 554-564 in [Code](#).

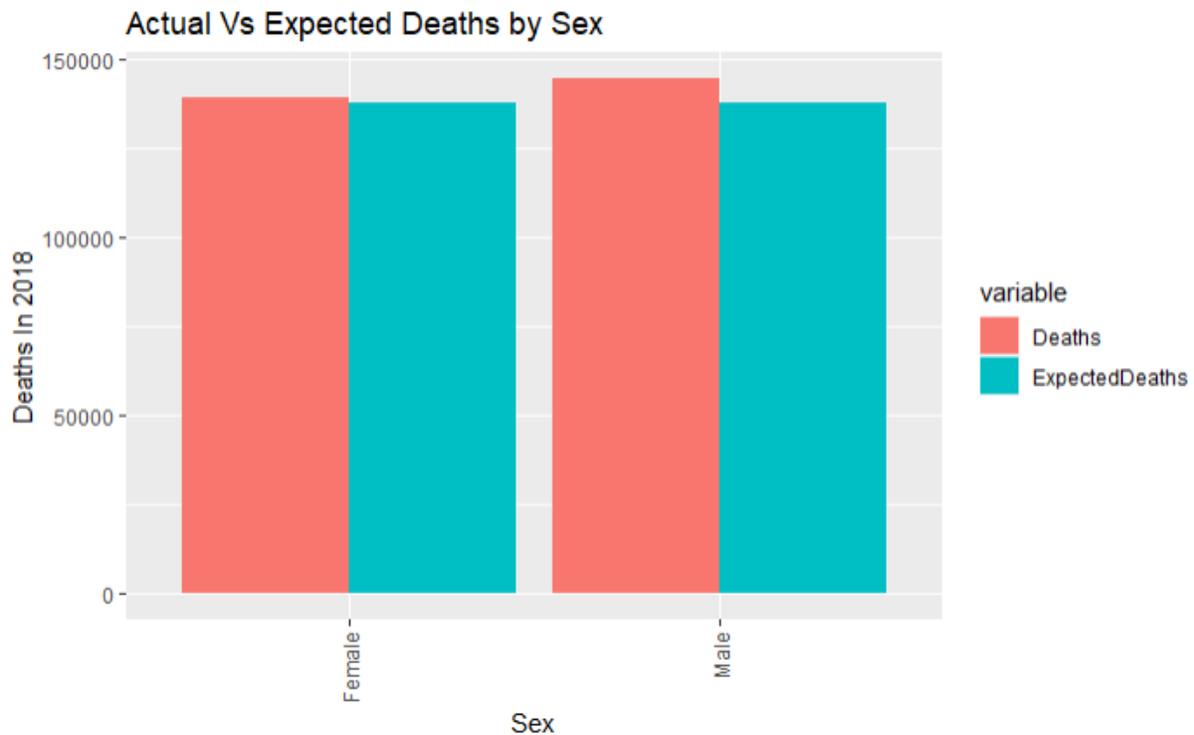


Figure. 15. Actual vs. Expected Deaths in Canada in 2018 by Sex, using linear models. Lines 569-579 in [Code](#).

From figures 12 and 13, the discrepancy between the predicted amount of deaths and actual amount of deaths in 2018 are consistent across the location and age categories. From [figure 14](#), the model did not perform as well for people that are single. From [figure 15](#), predictions are more accurate for females than males. Below are some graphs that display the accuracy of the linear model for specific risk classes.

e.g., the risk class for males aged 15-44, from Ontario and Single.

RED - Actual Deaths 2018, **BLUE** - Predicted Deaths 2018, **BLACK** - Historical Deaths from 1991-2017

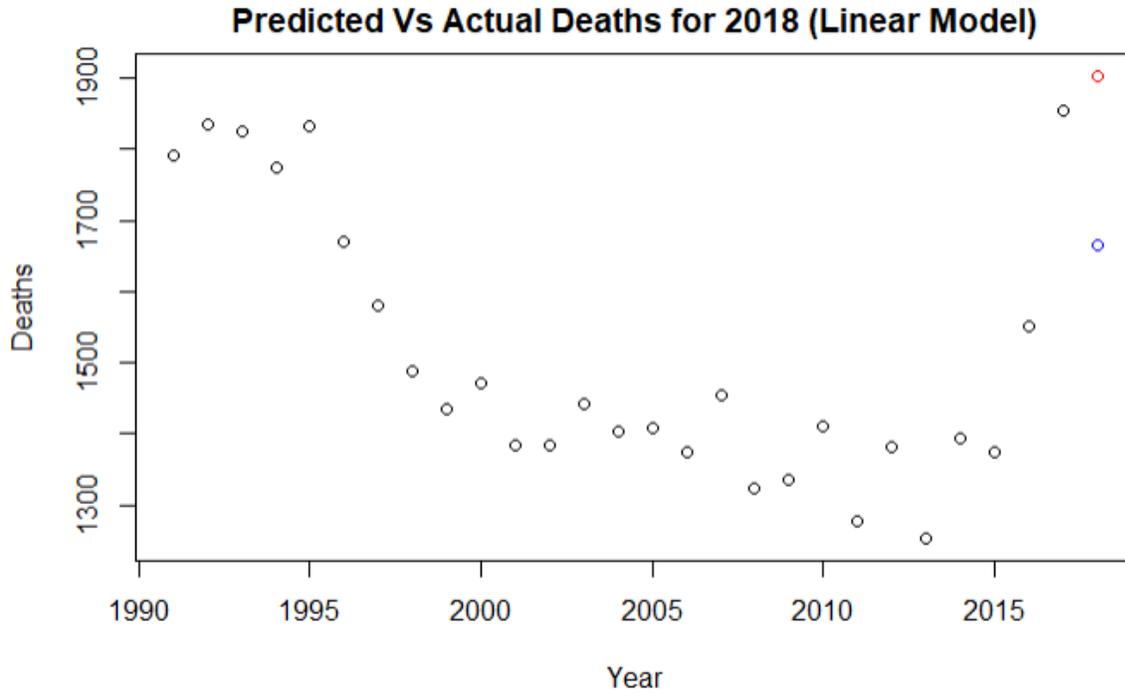


Figure. 16. Predicted amount of deaths vs. actual amount of deaths for males aged 15-44, from Ontario and Single, 2018. From linear model. Line 450 in [Code](#).

For this risk class the model performed poorly. The predicted amount of deaths is 1,665 and the actual amount of deaths is 1,903. This gives a percent error of 14.3%.

e.g., the risk class for females aged 85+, from Ontario and widowed.

RED - Actual Deaths 2018, **BLUE** - Predicted Deaths 2018, **BLACK** - Historical Deaths from 1991-2017

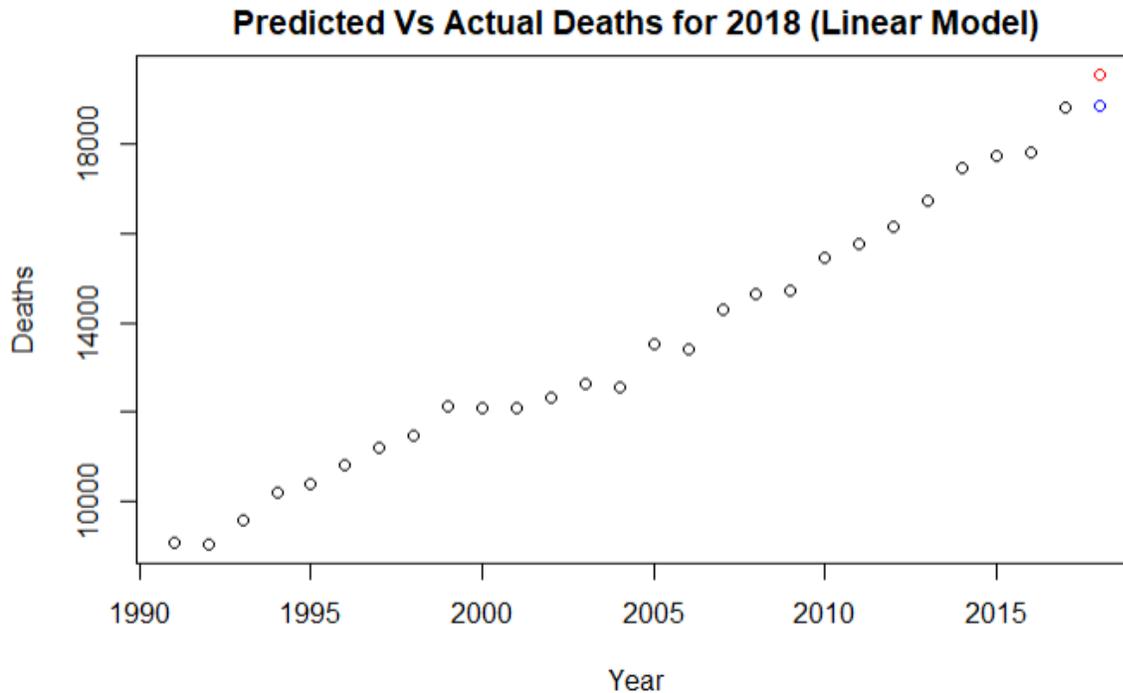


Figure. 17. Predicted amount of deaths vs. actual amount of deaths for females aged 85+, from Ontario and widowed, 2018. From linear model. Line 463 in [Code](#).

For this risk class the model performed well. The predicted amount of deaths is 18,865 and the actual amount of deaths is 19,564. This gives a percent error of 3.7%.

6.3. Gradient Boosting as a solution

The gradient boosting machine learning algorithm was applied next to the dataset. Once again the dataset was partitioned into its risk classes. The gradient boosting algorithm in R was applied to construct a unique model for each risk class that would predict the number of deaths in 2018. The total number of deaths for the year 2018 again is predicted as the sum of the predicted deaths for each risk class. That is:

$$\hat{y} = \sum_{c=1}^n \hat{D}_c$$

Since the gradient boosting machine algorithm requires many trees, it is computationally expensive, requiring a lot of memory and time. For this reason, to demonstrate its use, the algorithm was applied only to risk classes for people living in Ontario.

The mapping process of the model is:

$$g: \{A, G, S, M; S_p, Y, P\} \rightarrow h: \{S_p, Y, P\} \rightarrow \mathbb{Z}_{\geq 0}$$

where g is a function that takes the age, gender, sex and marital status and maps it to another function h ; h is a function produced by the gradient boosting algorithm. It takes the population, year and smoking percent and maps it to the number of deaths in the risk class.

6.3.1. Training the model

The training data was all data from 1991-2017. The model for every risk class was trained using the same sets as the linear model in section 6.2 .

Training data for risk class $C = \{x : A = a, G = g, S = s, M = m, 1991 \leq Y \leq 2017\}$

i.e., the training data for a risk class consists of observations with the same age, gender, sex, marital status and are from 2017 or earlier.

To train each model, the cross validation method was used to assess performance. After testing values for the number of folds from the set $\{3, 5, 8, 10, 12\}$, 10 folds seemed to give the best accuracy with a computation time of 25 seconds for each risk class.

A fold is another name for a split of the data. 10-fold cross-validation is training on 90% of the data and validating on the other 10%; 9 folds make up the training set and 1 fold makes up the validation set. It repeats the training and validation process 10 times by rotating the folds through the training set and the validation set so that there are different results each time. The average of the errors from the 10 trials is computed.

6.3.2. Results

Table 7. Comparing actual and predicted deaths in Ontario in 2018 using Gradient Boosting

Ontario

Year	Actual Deaths	Expected Deaths	Percent Error
2018	106,991	101,882	5.1%

Overall, Table 7 shows this model gave a prediction for the total number of deaths in Ontario for the year 2018 to be 101,882 whereas the actual number of deaths was 106,991. This is a percent error of 5.1%

The square root of the average squared deviations was 174.3.

6.3.3. Visualization on Results

Below are some visualizations that display the accuracy of the gradient boosting models across all the risk class categories.

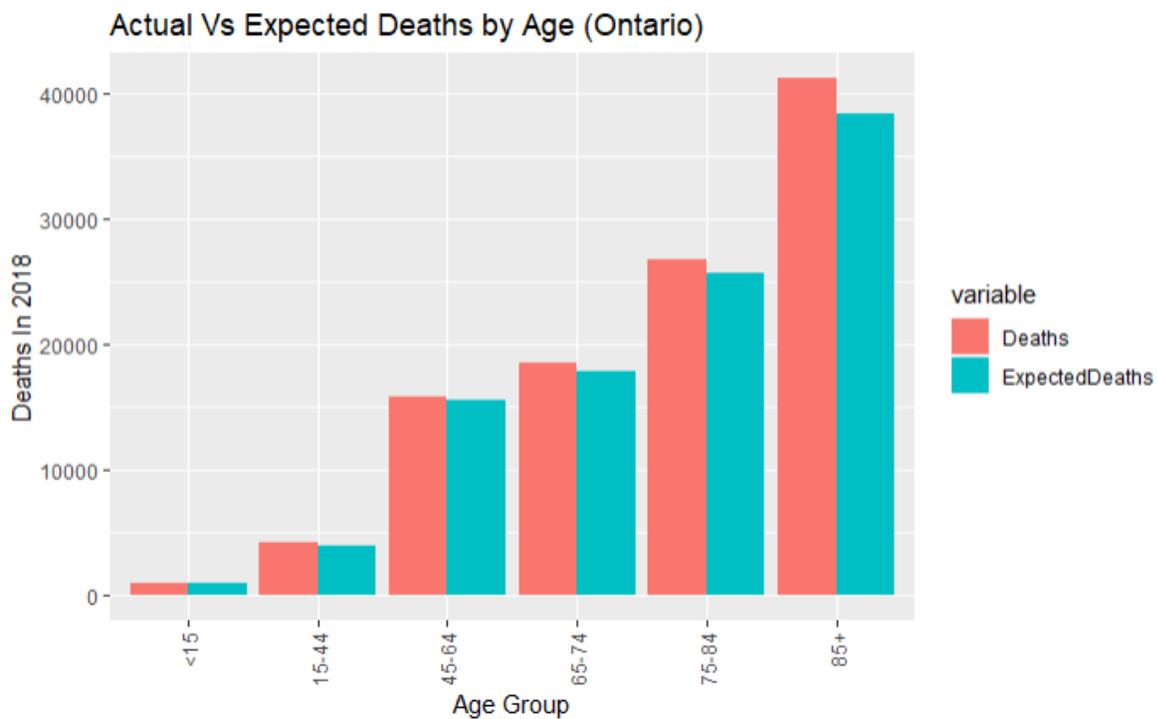


Figure 18. Actual vs. Expected deaths in Ontario in 2018, using gradient boosting. Lines 707-717 in [Code](#).

Figure 18 shows the gradient boosting method did a poor job predicting the number of deaths for people in the age group 85+. Since a large proportion of the total number of deaths in Canada come from people 85 or older, the large discrepancy in the actual and expected number of deaths among people 85+ has a significant negative impact on the accuracy of the model. This difference explains from where most of the percent error of 5.1% comes.

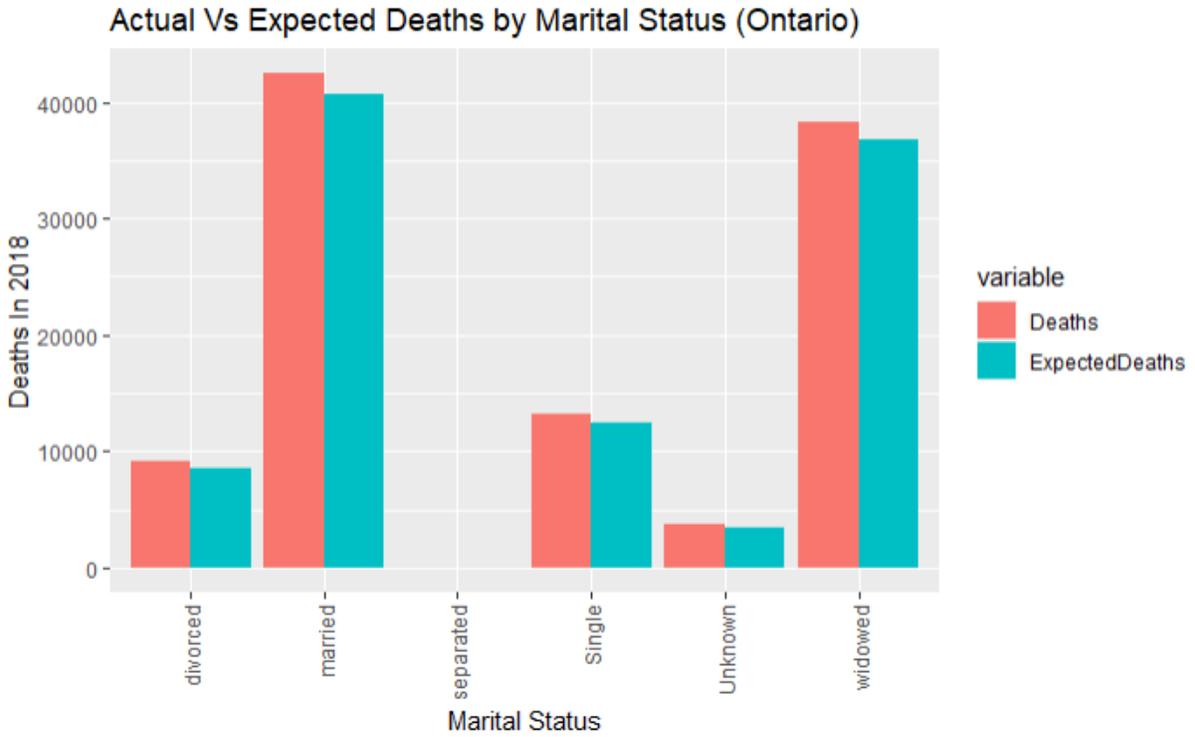


Figure 19. Actual vs. Expected deaths in Ontario in 2018 by marital status, using gradient boosting. Lines 722-732 in [Code](#).

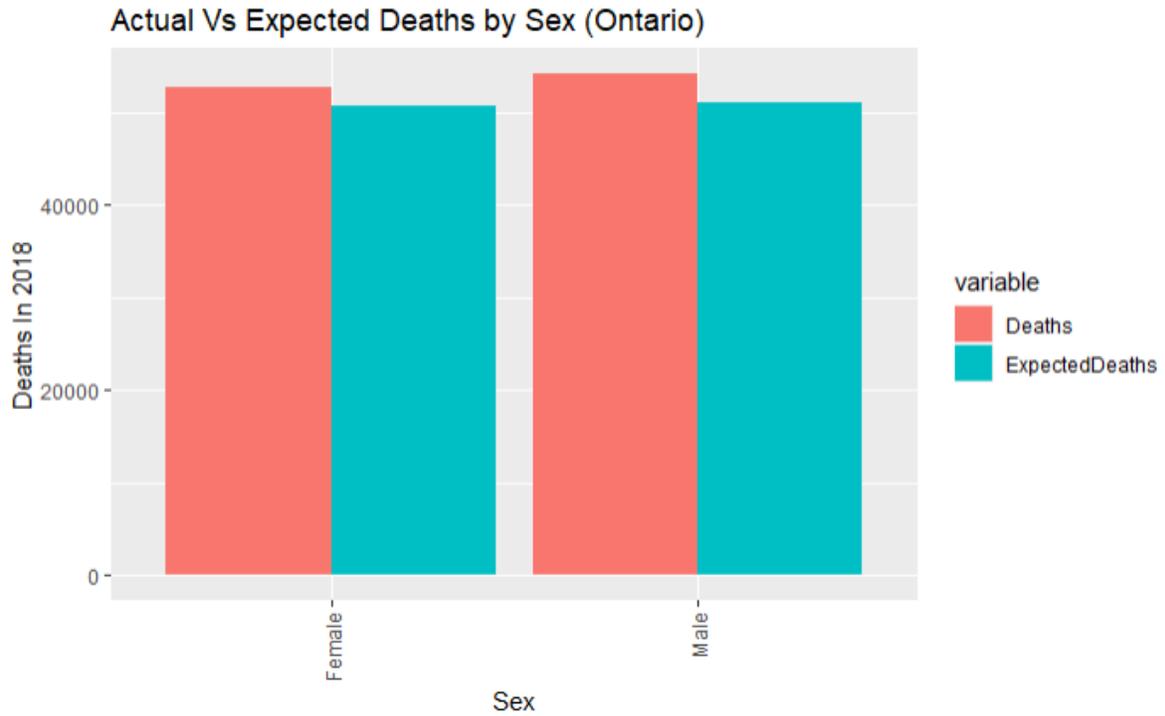


Figure 20. Actual vs. Expected deaths in Ontario in 2018 by sex, using gradient boosting. Lines 737-747 in [Code](#).

In figures 18-20, results are consistent across all three categories age, marital status and sex. Percentage error for each of the six marital statuses are roughly the same as shown in [figure 19](#). Percentage error for females and males are also similar with females being slightly less as shown in [figure 20](#).

Below are some graphs that display the accuracy of the gradient boosting model for specific risk classes.

e.g., the risk class for males aged 15-44, from Ontario and Single.

RED - Actual Deaths 2018, **BLUE** - Predicted Deaths 2018, **BLACK** - Historical Deaths from 1991-2017

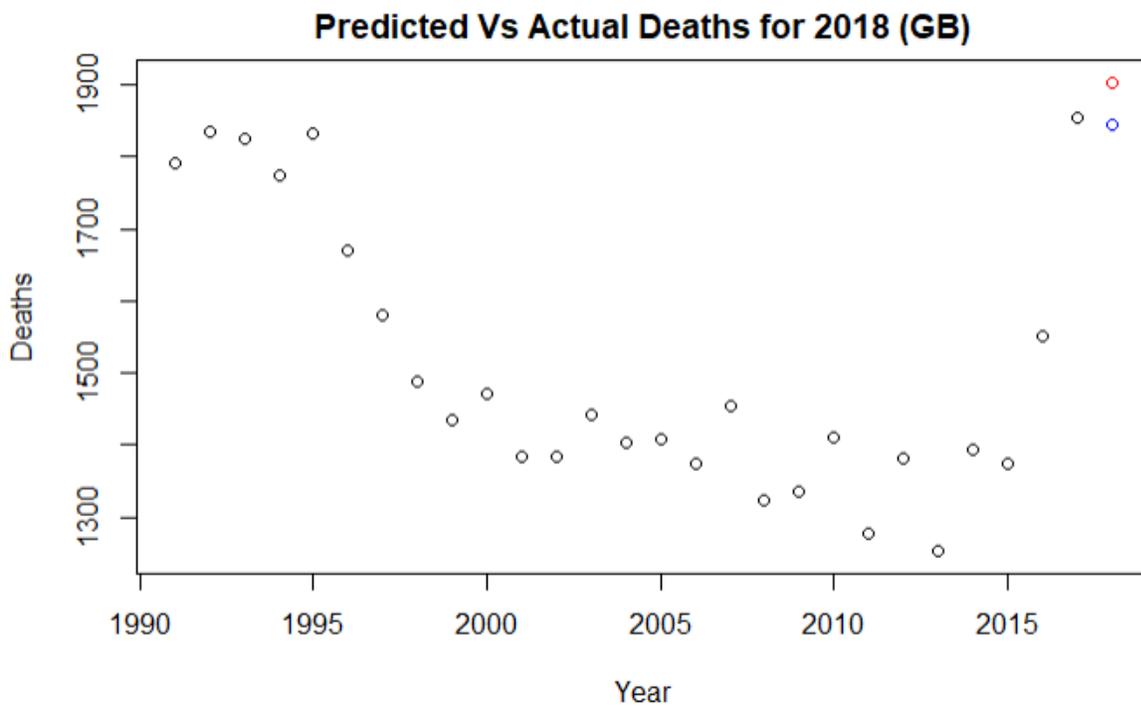


Figure. 21. Predicted amount of deaths vs. actual amount of deaths for males aged 15-44, from Ontario and Single, 2018. From gradient boosting model. Line 776 in [Code](#).

The Gradient Boost model had a much better prediction than the linear model for the number of deaths in the risk class of males aged 15-44 from Ontario and Single. The predicted amount of deaths is 1,844 and the actual amount of deaths is 1,903. This gives a percent error of 3.2%.

e.g., the risk class for females aged 85+, from Ontario and widowed.

RED - Actual Deaths 2018, **BLUE** - Predicted Deaths 2018, **BLACK** - Historical Deaths from 1991-2017

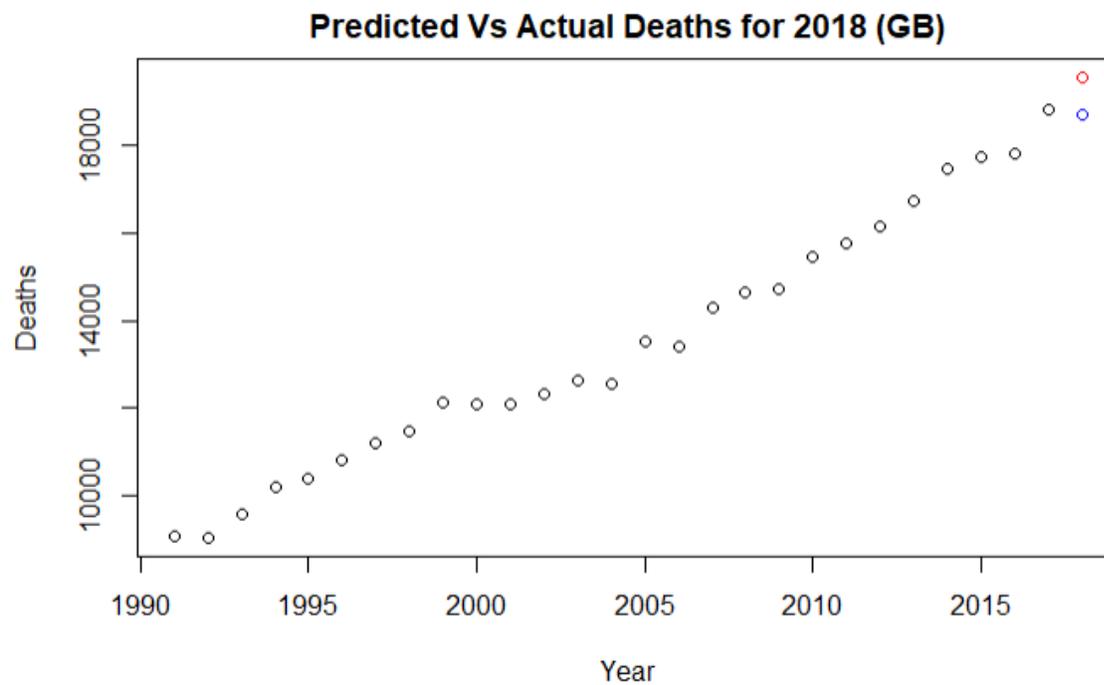


Figure. 22. Predicted amount of deaths vs. actual amount of deaths for females aged 85+, from Ontario and widowed, 2018. From gradient boosting model. Line 781 in [Code](#).

The Gradient Boost model had a slightly worse prediction than the linear model for the number of deaths in the risk class of females aged 85+, from Ontario and widowed. The predicted amount of deaths is 18,719 and the actual amount of deaths is 19,564. This gives a percent error of 4.5%.

6.4. Support Vector Machines Regression as a solution

Lastly, the support vector machines algorithm was applied to the dataset. Once again the dataset was partitioned into its risk classes. The support vector machines algorithm in R was applied to construct a unique model for each risk class that would predict the number of deaths in 2018. The total number of deaths for the year 2018 is predicted as the sum of the predicted deaths for each risk class. That is:

$$\hat{y} = \sum_{c=1}^n \hat{D}_c$$

where \hat{y} is the total expected number of deaths and \hat{D}_c is the expected number of deaths in risk class c . \hat{D}_c is estimated by fitting a support vector regression model.

$$\widehat{D}_c = h_c: \{S_p, Y, P\} \rightarrow \mathbb{Z}_{\geq 0}$$

6.4.1. Training the model

The models for each risk class c were fit using the historical results from 1991-2017. The training data is partitioned the exact same way as the other two previous models with training data for risk class

$$C = \{x : A = a, G = g, S = s, M = m, 1991 \leq Y \leq 2017\}$$

i.e., the training data for a risk class consists of observations with the same age, gender, sex, marital status and are from 2017 or earlier.

6.4.2. Predictions

Table 8. Comparing actual and predicted deaths in Canada in 2018 using SVM

Canada

Year	Actual Deaths	Expected Deaths	Percent Error
2018	283706	265446.6	6.9%

The Support vector machine model gave a prediction for the total number of deaths in Canada for the year 2018 to be 265,447. The actual number of deaths was 283,706. This is a percent error of 6.9%. The square root of the average squared deviations was 105.9.

Table 9. Comparing actual and predicted deaths in Ontario in 2018 using SVM

Ontario

Year	Actual Deaths	Expected Deaths	Percent Error
2018	106991	98831.64	8.3%

The prediction for the total number of deaths in Ontario for the year 2018 is 98,832. The actual number of deaths was 106,991. This is a percent error of 8.3%. The square root of the average squared deviations is 298.4 .

6.4.3. Visualizations

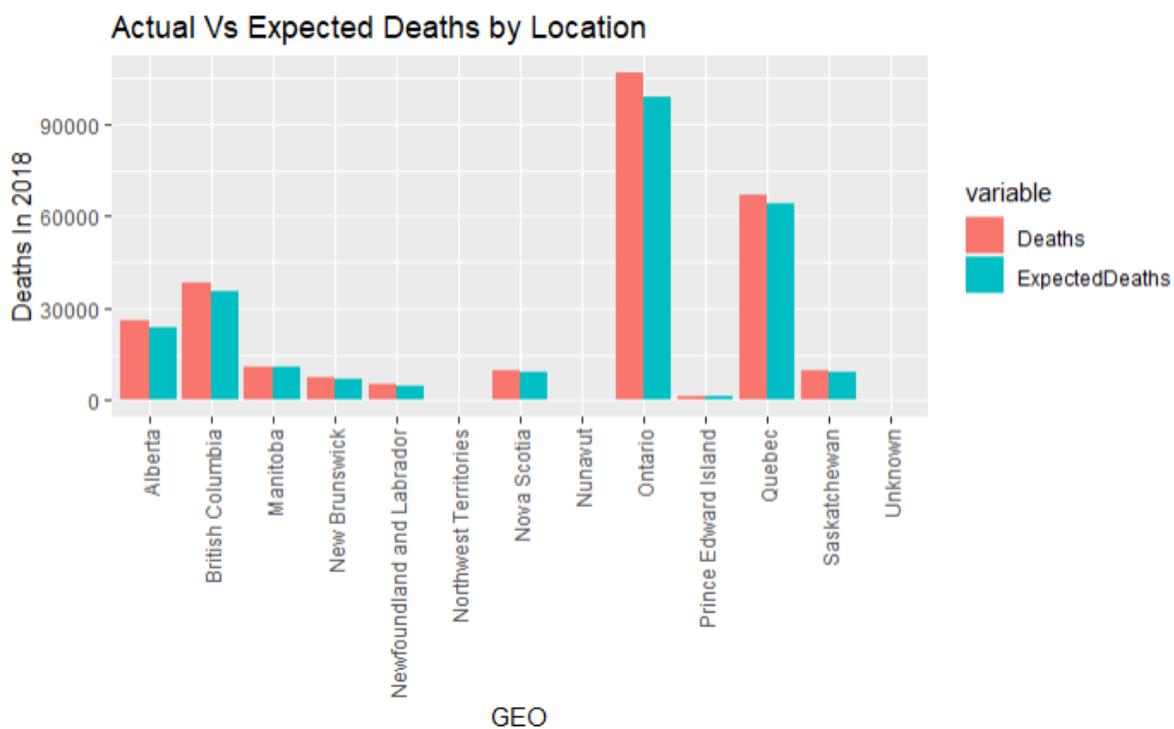


Figure 23. Actual vs expected deaths in Canada in 2018 by geographical region, using SVM. Lines 923-933 in [Code](#).

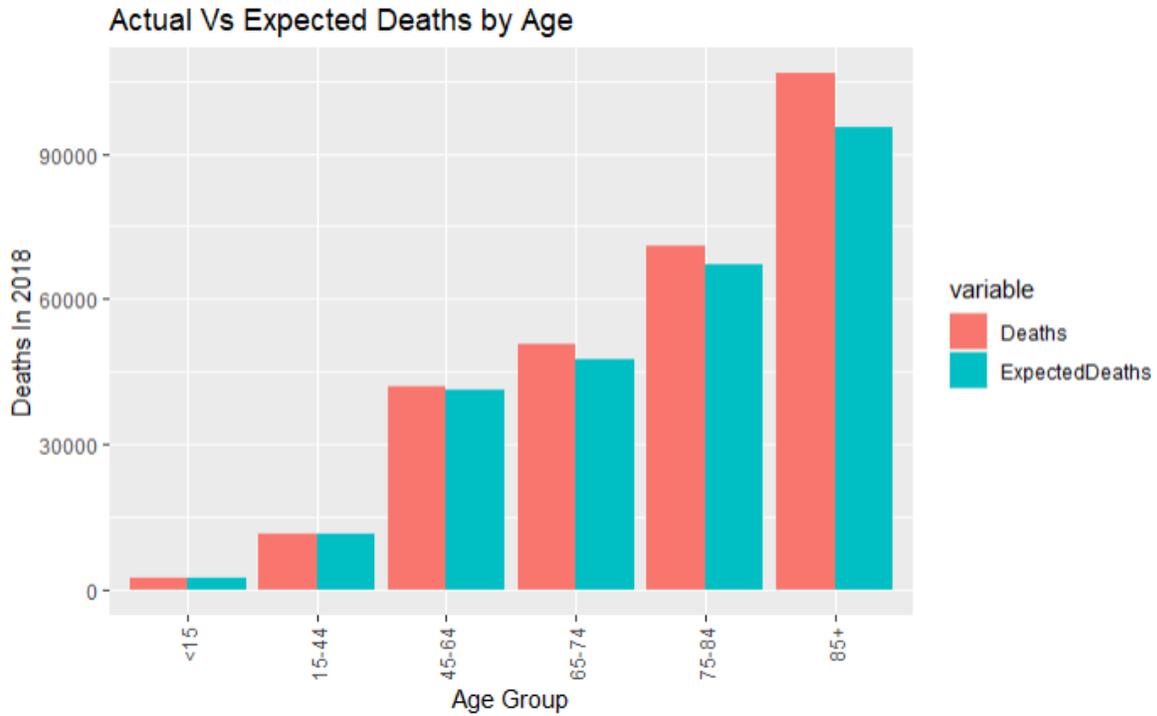


Figure 24. Actual vs expected deaths in Canada in 2018 by age group, using SVM. Lines 938-948 in [Code](#).

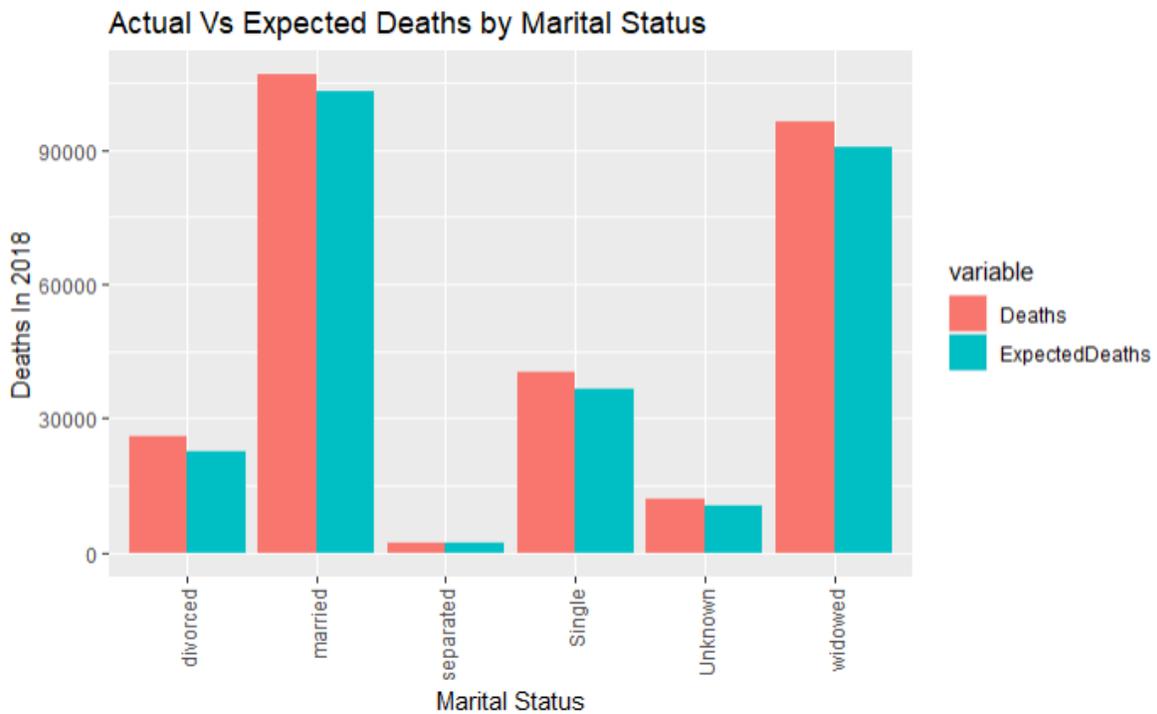


Figure 25. Actual vs expected deaths in Canada in 2018 by marital status, using SVM. Lines 953-963 in [Code](#).

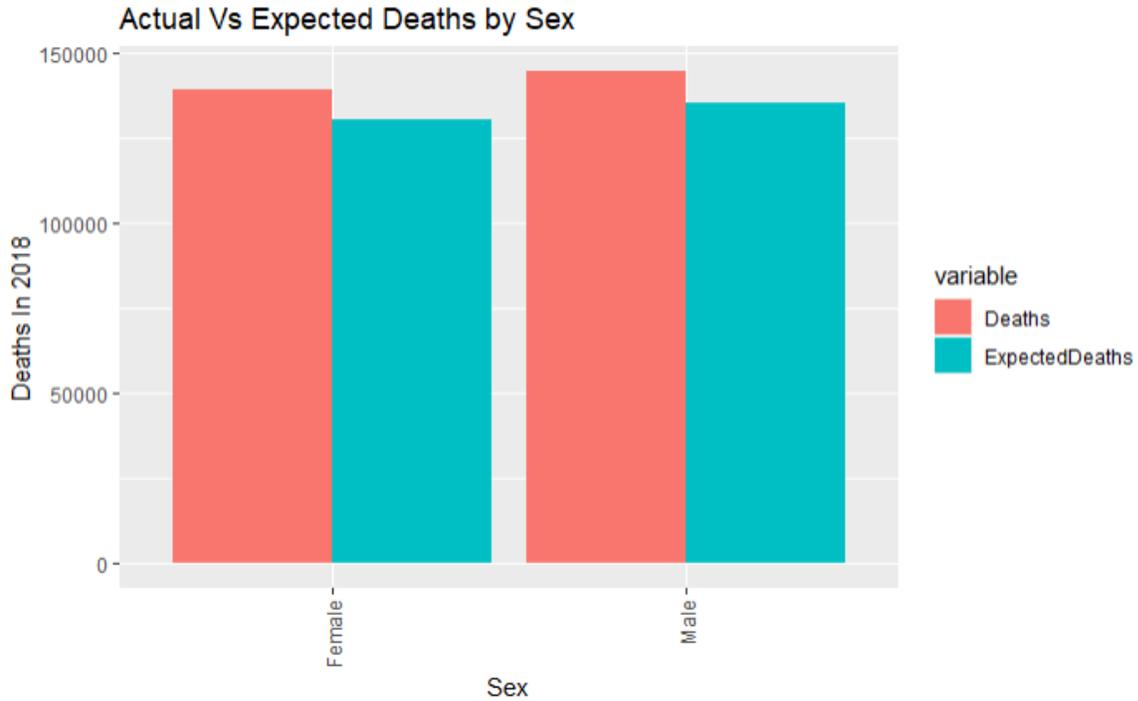


Figure 26. Actual vs expected deaths in Canada in 2018 by sex, using SVM. Lines 967-977 in [Code](#).

From figures 23-26, the accuracy of the predicted number of deaths are consistent across locations and sexes. The model did better at predicting the number of deaths of married people in Canada compared to the other marital statuses, as shown in figure 25. The model did worse at predicting the number of deaths of people age 85+ in Canada compared to the other age groups, as shown in figure 24. Discrepancies between the actual number of deaths and the expected number of deaths using SVM are clearly larger than those produced from the linear model and the gradient boosting model.

Below are some graphs that display the accuracy of the gradient boosting model for specific risk classes. e.g., the risk class for males aged 15-44, from Ontario and Single.

RED - Actual Deaths 2018, **BLUE** - Predicted Deaths 2018, **BLACK** - Historical Deaths from 1991-2017

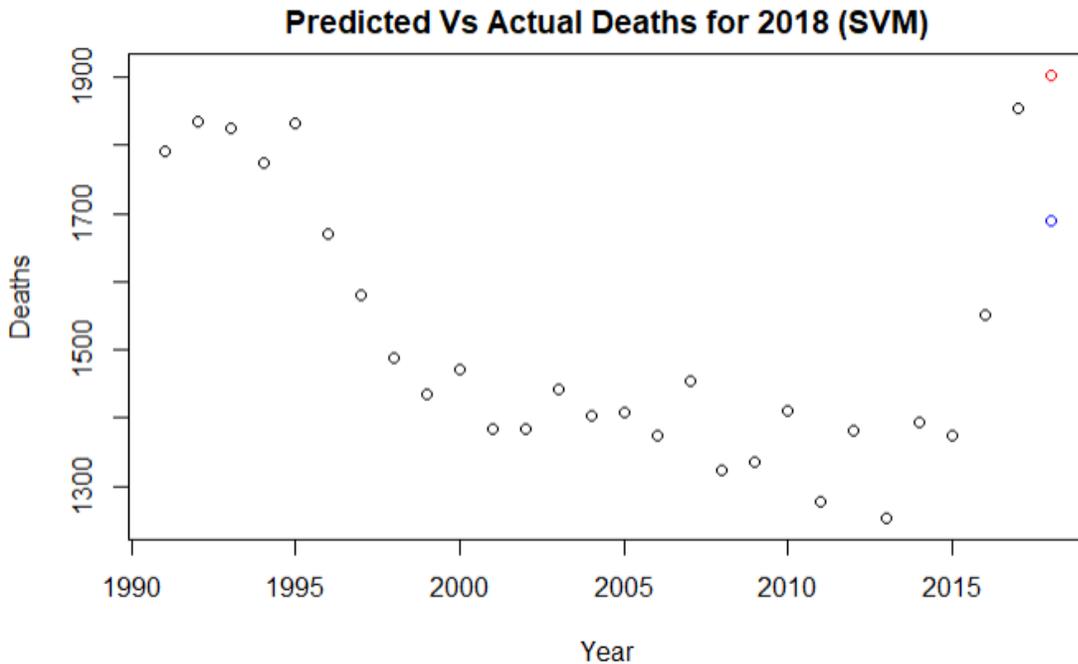


Figure. 27. Predicted amount of deaths vs. actual amount of deaths for males aged 15-44, from Ontario and Single, 2018. From SVM. Line 912 in [Code](#).

SVM had the second best prediction out of the three models for the number of deaths in the risk class of males aged 15-44 from Ontario and single. The predicted amount of deaths is 1689 and the actual amount of deaths is 1,903. This gives a percent error of 12.7%.

e.g., the risk class for males aged 15-44, from Ontario and Single.

RED - Actual Deaths 2018, **BLUE** - Predicted Deaths 2018, **BLACK** - Historical Deaths from 1991-2017

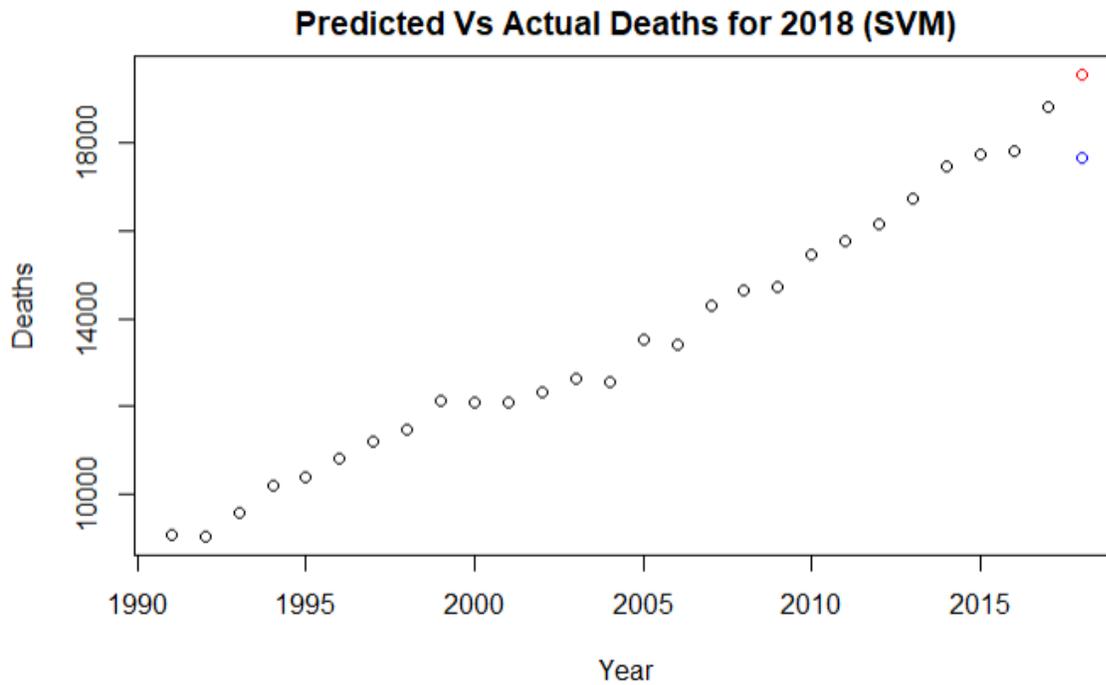


Figure. 28. Predicted amount of deaths vs. actual amount of deaths for females aged 85+, from Ontario and widowed, 2018. From SVM. Line 916 in [Code](#).

SVM had the worst prediction out of the three models for the number of deaths in the risk class of females aged 85+, from Ontario and widowed. The predicted amount of deaths is 17,679 and the actual amount of deaths is 19,564. This gives a percent error of 10.7%.

6.5. Summary of Results

Table 10. Summary of actual and predicted deaths (Canada and Ontario, 2018) using linear model, gradient boosting, and support vector machine models

Model	Location	Actual Deaths	Expected Deaths	RMSE
Linear	Canada	283,706	275,405.7	58.2
Linear	Ontario	106,991	102,296.8	156.8
Gradient Boost	Canada	283,706	NA	NA
Gradient Boost	Ontario	106,991	101,881.8	174.3
SVM	Canada	283,706	265,446.6	105.9
SVM	Ontario	106,991	98,831.6	298.4

Table 10 above shows that the best machine learning algorithm was the linear model. The linear model gave the closest predictions to the actual number of deaths in 2018 for both Canada and Ontario. It also had the best RMSE, which is the square root of the average deviations and is measured as:

$$RMSE = \sqrt{\frac{\sum_{c=1}^N (D_c - \hat{D}_c)^2}{N}}$$

7. Constructing Mortality Tables

The results from one of the models can be used to construct mortality tables as follows:

1. Let state 0 be the age group for 15 and younger (≤ 15)

Let state 1 be the age group for 15-44 year-olds

Let state 2 be the age group for 45-64 year-olds

Let state 3 be the age group for 65-74 year-olds

Let state 4 be the age group for 75-84 year-olds

Let state 5 be the age group for 85 and older (85+)

2. Fix the variables sex, marital status and geo so that $S = s$, $M_s = m$ and $G = g$.

3. Denote the population of state x as l_x .

4. Then the construction of a mortality table can be done as per Table 11 below for all the people that share this same sex, marital status and geo.

Table 11. Example of calculations for mortality table

State (x)	l_x	$l_x' = l_x - d_x$	d_x	p_x
0	l_0	$l_0 - \widehat{D}_0$	\widehat{D}_0	$\frac{\widehat{D}_0}{l_0}$
1	l_1	$l_1 - \widehat{D}_1$	\widehat{D}_1	$\frac{\widehat{D}_1}{l_1}$
2	l_2	$l_2 - \widehat{D}_2$	\widehat{D}_2	$\frac{\widehat{D}_2}{l_2}$
3	l_3	$l_3 - \widehat{D}_3$	\widehat{D}_3	$\frac{\widehat{D}_3}{l_3}$
4	l_4	$l_4 - \widehat{D}_4$	\widehat{D}_4	$\frac{\widehat{D}_4}{l_4}$
5	l_5	$l_5 - \widehat{D}_5$	\widehat{D}_5	$\frac{\widehat{D}_5}{l_5}$

where d_x , the number of deaths in the next year, is produced using the results of the model and l_x' is the number of Canadians in state x in the next year.

Note: l_x is a decreasing sequence. Immigration and births are not considered. In actuarial terms, the table does not adjust for any new business. It is only for the policies currently in force.

References

Chapter 1 - Introduction

[1] Susanna Levantesi and Virginia Pizzorusso. (2019). "Application of Machine Learning to Mortality Modeling and Forecasting". <https://www.mdpi.com/2227-9091/7/1/26>

Chapter 2 - Mortality Experience for Life Insurers

[2] Itre Mtalai, (2021), STAT3661 A Life Contingent Risk modelling I Winter 2021, Retrieved from https://culearn.carleton.ca/moodle/pluginfile.php/4373851/mod_resource/content/6/Ch2_stat3661A_2021.pdf, February 2021

Chapter 3 – Machine Learning Algorithms for Regression

3.1 Linear Model

[3] Jason Browlee, (2016), "Linear Regression for Machine Learning", Retrieved From <https://machinelearningmastery.com/linear-regression-for-machine-learning/>, March 2021

3.2 Gradient Boosting Decision Tree

3.2.2 Random forests (RF)

[4] Leo Breiman and Adele Cutler, (2004), "Random Forests", Retrieved from https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm, February 2021

3.2.3

[5] Bradley C. Boehmke, (2018), "Gradient Boosting Machines", Retrieved from http://uc-r.github.io/gbm_regression, March 2021

3.3 Support Vector Regression Model

[6] UBC, (2008), "Lecture 9: Support Vector Machines", Retrieved From https://www.cs.ubc.ca/~arnaud/cs540/lec9_CS540_handouts.pdf, February 2021

Chapter 4 - Introducing the Data

[7] School of Public Health and Health Systems university of Waterloo, (2017), "Smoking in Canada", retrieved from <https://uwaterloo.ca/tobacco-use-canada/adult-tobacco-use/smoking-canada>, February 2021

Chapter 5 - Data Visualizations

Chapter 6 - Developing Machine Learning Models

6.3.1 Training the Gradient Boosting Model

[8] STHDA, (2018), “Gradient Boosting Essentials in R Using XGBoost”, Retrieved from <http://www.sthda.com/english/articles/35-statistical-machine-learning-essentials/139-gradient-boosting-essentials-in-r-using-xgboost/#:~:text=Stochastic%20gradient%20boosting%2C%20implemented%20in,and%20columns%20in%20each%20round.&text=Boosting%20can%20be%20used%20for%20both%20classification%20and%20regression%20problems.> , March 2021

[9] UBC, (2017), “CPSC 340: Machine Learning and Data Mining ”, Retrieved from <https://www.cs.ubc.ca/~schmidtm/Courses/340-F17/L5.pdf>, March 2021

[5] Bradley C. Boehmke, (2018), “Gradient Boosting Machines”, Retrieved from http://uc-r.github.io/gbm_regression, March 2021

6.4.1 Training the SVM

[10] Chaitanya Sagar, (2017), “Building Regression Models in R using Support vector Regression”, Retrieved from <https://www.kdnuggets.com/2017/03/building-regression-models-support-vector-regression.html>, March 2021

Code

https://github.com/cameronmcmullen/CanadianMortalityStudy/blob/main/CanadianMortalityStudy_Code.Rmd