



An Innovative Flood Prediction System Using Improved Machine Learning Approach

Cynthia Cui and Leonardo Cui

Ages 16 and 18 | Fredericton, New Brunswick

Canada-Wide Science Fair 2019 Finalists | Intact Climate Change Resilience Award 2019

Grand Award Winner: River Valley Regional Science Fair 2019

Flood is a highly chaotic, strongly emergent, and naturally occurring pattern. Flood prediction is important for building resilience to natural disasters by proactively managing and minimizing the impacts of natural hazards on our everyday lives. This project attempts to create a model that can accurately and efficiently map water level as a function of climate conditions to forecast flood levels at a given point in the future. Machine learning, as one of the Artificial Intelligence methods, contributes highly to the advancements of prediction systems to provide better performance and cost-effective solutions. Most of the current flood prediction models are primarily concerned with heavy rainfall and hurricanes. However, flooding in New Brunswick, commonly referred to as the spring flood, is primarily caused by rapid snowmelt, heavy rainfall and ice jams occurring during the spring (Figures 1 and 2). To create an accurate prediction, we implemented machine learning to discover a new prediction model for this type of flooding, using local environmental data and considering different hydrology and climate variables.

HYPOTHESIS

We believed that local climate data such as temperature and precipitation, including rainfall and snow on the ground, will have significant effects on water level during the flood season.

METHODS AND PROCEDURE

Data Collection

Data was collected on water levels of the Saint John River, daily average temperature, min/max daily temperature, and daily rainfall amount and snow on the ground in New Brunswick. The data came from New Brunswick River Watch, Canadian Centre for Climate Services and WeatherStats which provide historical weather data from Environment and Climate Change Canada. We clipped the data for the time frame that was relevant to our model.

Procedure

A multiple linear regression model was chosen for machine learning as we thought it would be the best option (Figure 3). We used historical water level during flood season as the dependent variable and a combination of climate data as explanatory variables to build a prediction model. We also considered lagged values such as yesterday's temperature. Our dataset was divided so two-thirds of the data was used for training and one-third for testing (Figure 4). We then trained the model with our training data and tested its performance with the testing data. We first evaluated models by comparing the R² values for each of our models. We then tested to see if our hypothesis was true on the coefficient parameters of each explanatory variable. Integration testing was used to evaluate

how the explanatory variables work together. Our project is implemented with Python and its Scikit-learn package, a powerful tool for machine learning. See Figure 5 for the machine learning process.

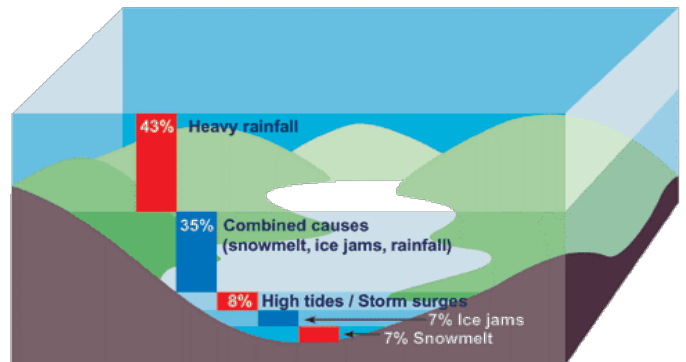


Figure 1. Causes of flooding in New Brunswick.



Figure 2. Snowmelt in New Brunswick.



This work is licensed under:
<https://creativecommons.org/licenses/by-nc-sa/4.0>
doi: 10.18192/csfj.v2i32020115119

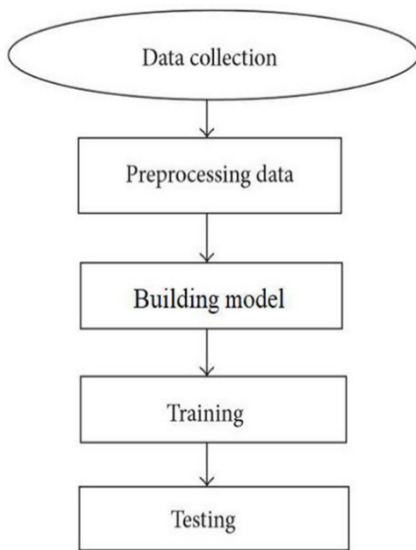


Figure 3. Data collection model.

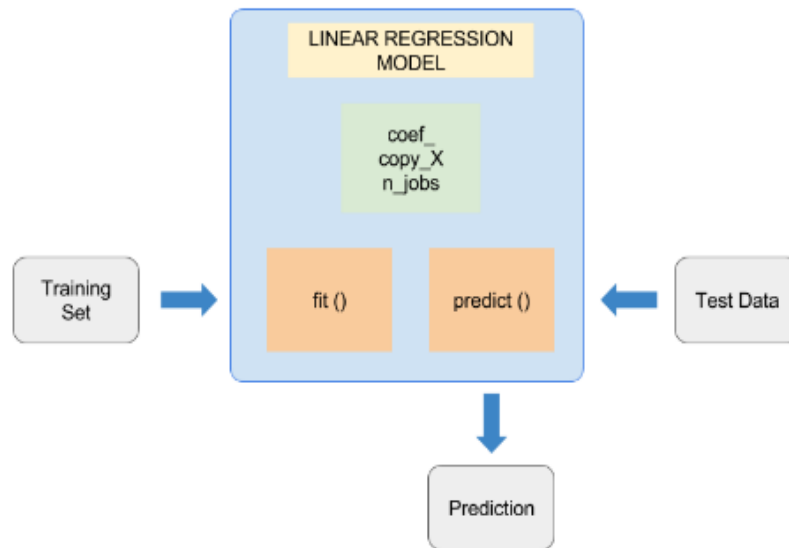


Figure 4. Linear regression model.

The Machine Learning Process

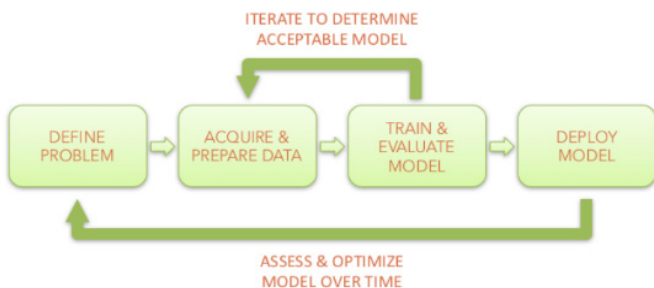


Figure 5. Machine learning process.

RESULTS AND OBSERVATIONS

We focused on the R-squared value and the coefficient values when looking at our results. The R-squared value shows the explanatory power of our regression model. The coefficients are important as they are what were used to construct the equation for our function, furthermore, we looked at the P value to assess the significance of the coefficients (Figure 6).

According to coefficients of the different variables in the observation table, we can write our linear model as:

$$y = 4.7577 + 0.0526 \times \text{Min Temperature} + 0.0526 \times \text{Yesterday Temperature} + 0.219 \times \text{Percipitation} + 0.0638 \times \text{Snow on Ground}$$

For example, when overnight Minimum Temperature is 1 degree and Yesterday Temperature is 15 degrees with Precipitation is 10 mm and snow on ground 4cm, the expected water level is 6.057 meters.

The R-square can be written as

$$R^2 = 1 - \left(\frac{\text{residual sum of squares}}{\text{total variance in } Y} \right)$$

It measures proportion of variability in Y that is explained by X using our model. Therefore, in our model, 63.0% of the variability in Y can be explained using X.

DISCUSSION

According to the goodness of the fit, R-squared shows our model can explain 63.0% of change of water level. According to hypothesis testing on the coefficient of independent variables, each explanatory variable - Minimum Temperature, Yesterday Temperature, Rainfall and Snow on Ground - have a significant effect on the change of water level. Most academic research and literature on flood prediction are limited to rainfall and hurricanes. However, for New Brunswick and much of Canada, floods are caused by rapid snowmelt, heavy rainfall, and ice jams. Spring is the peak flood season as snowmelt increases runoff and ice jams occur. Temperature is an interesting factor as high temperatures may cause more snowmelt but low temperatures may slow down the snowmelt and hold water in the soil. Considering the local climate and geographic factors, our new prediction model will



provide more accurate water level prediction and can be applied for areas like New Brunswick. There are three contributions from our research:

We redesigned the explanatory variables and trained our model with a local environment dataset. We added “Temperature” and “Snow on Ground” as an explanatory variable to represent the effect of snowmelt.

Geographic features like forests and soil type may also have an effect and cause a delay between when rainfall occurs and when the water level increases. We hope to include the lag time of precipitation into our new model.

We discovered that the lowest temperature may play opposite effects on the water level. Freezing temperature overnight may make snow melt gradually. The minimum daily temperature we added to the model does show a significant effect on the change of water level.

APPLICATION

A new approach of data visualization was implemented with a web-based simulation to clearly illustrate flooding zones using gradual colour change. With Google Earth Engine API which combines a multi-petabyte catalogue of satellite imagery and geo-spatial data such as Canadian Digital Elevation Model (CDEM), we can visualize and simulate the change according to the change in water level. This data visualization will contribute to risk reduction, policy suggestion, and reduction of property damage associated with floods.

We recommended this to NB Power, NB EMO, and the local community. These groups can predict and simulate the effect of different water levels for emergency response planning. The user is able to use our web-based simulation in either map view or satellite.

CONCLUSION

After applying machine learning and evaluating different prediction models, our hypothesis was proven to be true. Local climate data has significant effects on water level during the flood season. Machine learning with local climate data can build a good quantitative model to predict water level during flood season.

Based on our research, we implemented a new approach of data visualization with a web-based simulation map to clearly illustrate flooding zones to help the local community.

FUTURE WORK

For future work, we plan to consider more explanatory variables such as the humidity of the air, tides, precipitation rate, and accumulated snowfall amount. We’d also like to see how the Upper St. John River affects floods in Fredericton. NB Power suggested that we include tides as an independent variable for our model to increase our model’s accuracy and we’re currently doing so by looking at the Lu-

OLS Regression Results

Dep. Variable:	LEVEL	R-squared:	0.630
Model:	OLS	Adj. R-squared:	0.625
Method:	Least Squares	F-statistic:	110.9
Date:	Mon, 08 Apr 2019	Prob (F-statistic):	6.36e-42
Time:	14:00:48	Log-Likelihood:	-106.32
No. Observations:	199	AIC:	220.6
Df Residuals:	195	BIC:	233.8
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.7577	0.076	62.484	0.000	4.608	4.908
MIN_TEMP	0.0526	0.007	7.053	0.000	0.038	0.067
TEMP_1	0.0515	0.007	7.078	0.001	0.037	0.066
PRECIPITATION	0.0219	0.005	4.041	0.005	0.011	0.033
SNOW_ON_GROUND	0.0638	0.022	2.852	0.004	0.020	0.108

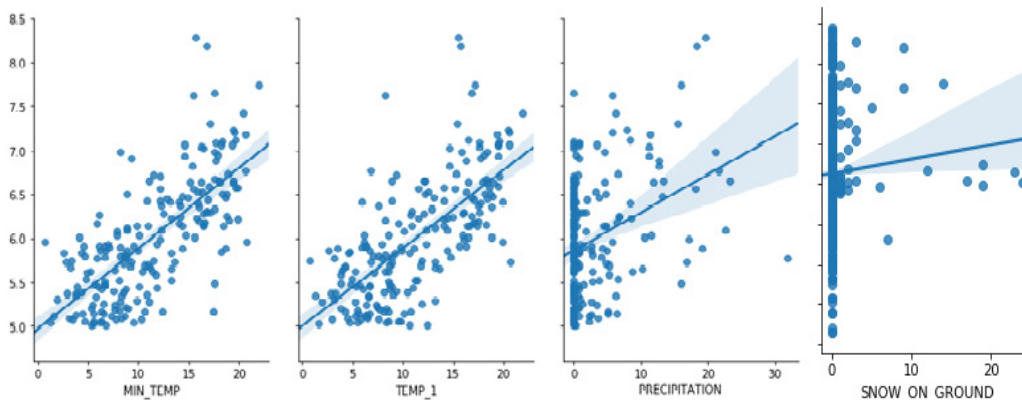


Figure 6. OLS regression results.

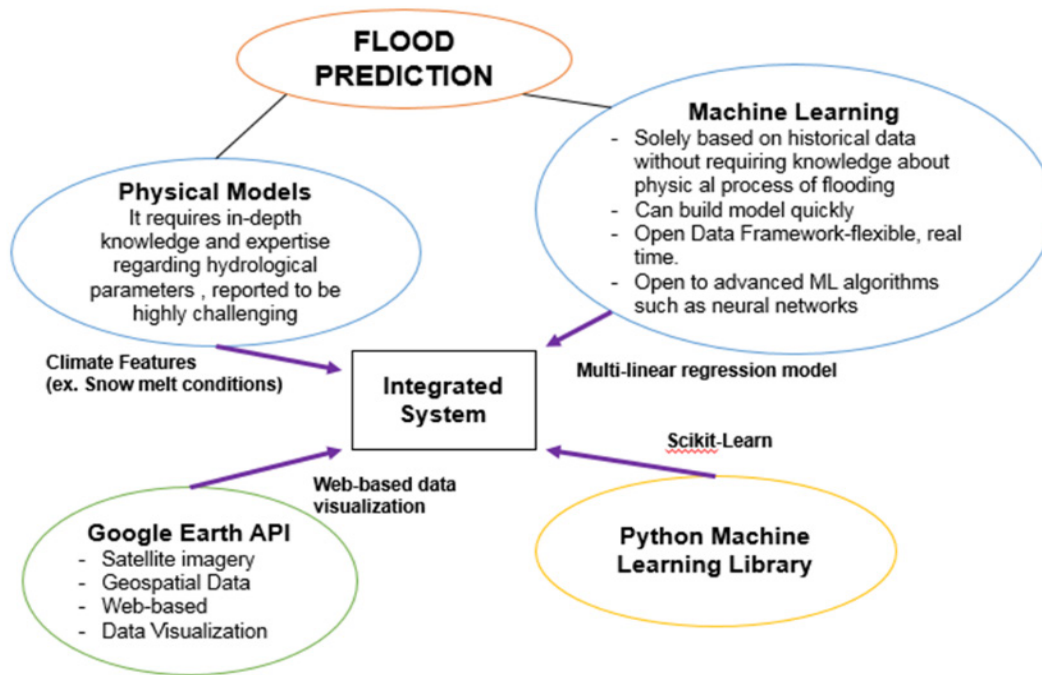


Figure 7. Integrated flood prediction system

nar calendar. We're close to creating a model that includes tides as an independent variable.

NB Power also mentioned the impact of the opening of the dam on flood levels, they mentioned in the following year, they'll consider using our model to assess what day to open the dam.

We're collecting historical data on the humidity of the air or data on the daily speed of wind. We're currently looking for data on precipitation rate. We're also collecting data from different stations to create more models for different rivers or different parts of the St. John River.

Furthermore, we plan on using a more complicated prediction model such as nonlinear regression or the Artificial Neural Network approach. We will also consider real-time data input from weather forecasts to our prediction system.

ACKNOWLEDGMENTS

We would like to express our gratitude to our teacher Mr. Clifford Cull as well as Mr. Pierre Lumsden, Leo's supervisor during his SHAD internship at NB Power, who both helped us in data collection and testing during flood season.

REFERENCES

Tsakiri, K.; Marsellos, A.; Kapetanakis, S. Artificial Neural Network and Multiple Linear Regression for Flood Prediction in Mohawk River, New York. *Water* 2018, 10, 1158

Duan, Y.; Liu, T.; Meng, F.; Luo, M.; Frankl, A.; De Maeyer, P.; Bao, A.; Kurban, A.; Feng, X. Inclusion of Modified Snow Melting and Flood Processes in the SWAT Model. *Water* 2018, 10,1715.

Anagnostopoulos, G. G., Koutsoyiannis, D., Christofides, A., Efstratiadis, A. & Mamassis, N. (2010) A comparison of local and aggregated climate model outputs with observed data. *Hydrological Sciences Journal*. Vol 55(7), 1094–1110

Mosavi, A.; Ozturk, P.; Chau, K.-W. Flood Prediction Using Machine Learning Models: Literature Review. *Water* 2018, 10, 1536.

Richard, C.; Gratton, D.J. The Importance of the Air Temperature Variable for the Snowmelt Runoff Modelling Using SRM. *Hydrological Processes* 2010 15(18): 3357-3370.

Kundzewicz, Z.W., et al., 2013. Flood risk and climate change: global and regional perspectives. *Hydrological Sciences Journal*, 59 (1), 1–28.

Government of New Brunswick: Flooding in New Brunswick <https://www2.gnb.ca/content/gnb/en/departments/elg/environment/content/flood.html>

Guide to training and deploying machine learning models using Python, <https://medium.freecodecamp.org/a-beginners-guide-to-training-and-deploying-machine-learning-models-using-python-48a313502e5a>

Scikit-learn: Linear Regression Python Example, https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html

DATA RESOURCES

Government of New Brunswick -Environment and Local Government, <https://www2.gnb.ca/content/gnb/en/departments/elg/environment/content/flood.html>

Canadian Centre for Climate Services, <https://www.canada.ca/en/environment-climate-change/services/climate-change/canadian-centre-climate-services.html>

Snow Depth - The Weather Network, <https://www.theweathernetwork.com/maps/snow>

Canadian Digital Elevation Model (CDEM), <https://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/download-directory->

Government of Canada: Climate data extraction tool, <https://climate-change.canada.ca/climate-data/#/Images Resource>

CTV Atlantic, <https://atlantic.ctvnews.ca/flood-risk-forces-evacuation-in-perth-andover-n-b-1.2334463>

GOOGLE'S NEW CLOUD AUTOML. <https://www.cloudpulsestrat.com/posts/googles-new-cloud-automl-broader-role-automated-machine-learning-ai>

Flooding in New Brunswick <https://www2.gnb.ca/content/gnb/en/departments/elg/environment/content/flood.html>

CBC New Brunswick, <https://www.cbc.ca/news/canada/new-brunswick/nb-flood-2019-april-28-1.5114372>, <https://www.cbc.ca/news/canada/new-brunswick/flooding-fredericton-rising-water-levels-1.5107001>



CYNTHIA CUI

Cynthia is a grade 11 student at Fredericton High School where she's actively involved through the student council, TEDxFrederictonHigh, and in athletics with swimming and badminton. In her free time, Cynthia rows and swims competitively and is a chess player who regularly represents Canada at international competitions. After being invited and attending a computer science workshop at the University of Waterloo, she was inspired to solve real-world problems with programming to change the world. Cynthia hopes her innovative approach can improve flood prediction models, specifically for areas with spring flooding. She hopes her model can benefit many people during flood season by helping minimize property damage and helping with risk reduction.

LEONARDO CUI

Leonardo is a grade 12 student at Fredericton High School. As a proud Frederictonian rower and resident, he has experienced both the beauty and vitalizing force of the Saint John River, and the chaotic and damaging effects each spring. He is a member of his school's Tech Crew and is interested in up and coming technology such as AI. In his free time, Leonardo rows and fences competitively co-chair the Fredericton Chess Club. He has represented Canada at international chess competitions and attended the SHAD program in 2018 at Queens, where the idea of flood prediction first occurred to him. An innovator at heart, Leonardo came up with a flood prediction approach that has the potential to solve real-life issues his community faces. After presenting at Flood Conversations Panels hosted by his city and receiving feedback from the general public, he realized the impact his model could minimize damage caused by flooding all the while preserving the natural beauty of his city."

