IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Predicting Floods with Deep Learning Neural Networks and Analyzing Their Socioeconomic Impacts with OLS Regressionn

Adrian Huihao Zhuang, Alex Huijie Zhuang

The King's School Parramatta, 87-129 Pennant Hills Road, North Parramatta, NSW 2151, Australia; zhuanghuihao2006@gmail.com
Mentor: Eric Sakk

ABSTRACT: The ability to predict flood occurrences and gauge their socio-economic impacts is highly essential for flood management. This research collated one of the largest datasets of historical floods (4000+ entries) ever used in Australian flood research and mapped flood locations to raster datasets of hydrological flooding factors created using ArcGIS software. The resulting dataset underwent unsupervised clustering to display statistical significance and was used to train a Deep Learning Neural Network with an 85-90% accuracy in predicting flood likelihoods given coordinate inputs. A K-nearest-neighbors optimisation algorithm was used to determine the most relevant hydrological factors in flood prediction. Seeking to address current research gaps in estimating the societal impacts of floods, we merged the historical floods dataset with Australian census data and performed Ordinary Least Squares regression, accounting for time & space fixed effects through categorical regression variables. We discovered that floods disproportionately impact Indigenous Australians as flooded areas see a disproportionate increase in Indigenous Australian population, hinting towards a dangerous possibility of environmental redlining. This research also quantified how floods affect select census statistics such as household income, unemployment, and migration. This research's regression model can be extended to gauge the impact of floods on any census statistic.

KEYWORDS: Earth and Environmental Sciences; Water Science; Flood Prediction; Neural Network; Regression; Socioeconomic Impacts.

Introduction

While many associate Australia as a sunburnt country with dry and arid sweeping plains, regions within Australia's east-coast States are susceptible to frequent flooding.[1] The Great Dividing Range blocks streams of moist air from the Tasman Sea, resulting in topographic rainfall and subsequent flooding in the States of Queensland, New South Wales, and Victoria.[2] Reports from the Australian government estimate that the heavy rain period of February to March of 2022 alone has caused a $5 billion hit to the Australian economy, and the Insurance Council of Australia (ICA) showed the insurance bill for storms and floods since January 2020 has reached $12.3 billion. This research aims to answer the question: "How can we predict the likelihood of a flood event and analyse its socio-economic repercussions?". This research has two major components:

1. Flood Prediction – Historical data on flooded and non-flooded regions in Queensland, Australia are mapped to a range of hydrological variables and used to train a deep learning neural network (DLNN). The DLNN is then used to assess the potential severity of previously unseen flood events.

2. Socio-Economic Inference – Historical flood data is organised into postal codes. Census data from flooded postal codes and non-flooded postal codes are analysed using ordinary least squares regression to determine the impacts of floods on sociological variables.

The significance and novelty of this research is that not only can we determine which geographical areas are prone to flooding through flood susceptibility modeling, we can also quantitatively estimate the societal impacts of flooding, in terms of housing market fluctuations, population changes, and even variations in local Indigenous Australian percentages. This result is valuable as floods are unpredictable exogenous events and natural experiments to analyse such societal impacts are extremely rare. Flood susceptibility modeling gives individuals more time to evacuate and emergency responders (State Emergency Services, Fire & Rescue, Paramedics) a forewarning to organise upcoming operations. Further, socio-economic inferences drawn from census data analysis will assist in better city planning, infrastructure development, and can provide a much-needed insight into redlining in Australia.[3]

Data Acquisition and Processing

All datasets used and created in this research fall within 10.7050° S to 28.1421° S latitude and 139.3556° E to 153.6165° E longitude. It is worth noting that a caveat with current research on flood analysis is its overly specific focus on local areas.[4-6] This research seeks to innovate upon existing research by conducting flood analysis across the entire State of Queensland – an area that is an order of magnitude larger than areas studied in existing literature – so that the conclusions obtained from this research are readily usable by governments for objectives of accurate forecast and identification of flood-prone and impacted areas, and first responder organizations to coordinate better rescue efforts.[7] Three datasets were prepared during this research:

DOI: 10.36838/v6i12.4

Data Used for Flood Prediction	Data Used for Socio-Economic Inference
1. Historical flood records 2. Hydrological flooding factors	1. Historical flood records 3. Australian census data

Figure 1: Datasets used for different sections of this research.

Historical flood records and hydrological flooding factors will be used to train a Deep-Learning Neural Network (DLNN) that acts as a binary classifier. Given a set of hydrological factors at the same geographical coordinate, the DLNN is able to determine whether a flood has occurred. On the other hand, Historical flood records can also be used with Australian census data to run Ordinary Least Squares (OLS) regression lines to ascertain the impacts of floods on census variables.

Important Variables In Each Dataset:

Important variables in Historical Flood Records dataset:
- Longitude
- Latitude

Important variables in Hydrological Flood Factors dataset:
- Elevation
- Slope
- Annual Rainfall
- Annual Mean Temperature
- 9am Humidity
- 3pm Humidity
- Stream Power Index (SPI)
- Sediment Transport Index (STI)
- Terrain Ruggedness Index (TRI)
- Topographic Wetness Index (TWI)

Important variables in Historical Flood Records:
- 'median_age',
- 'median_family_income',
- 'median_household_income',
- 'median_housing_monthly_loan_repayment',
- 'median_rent',
- 'average_household_size',
- 'indigenous_count_male',
- 'indigenous_count_female',
- 'population_male',
- 'population_female',
- 'indigenous_percentage_male',
- 'indigenous_percentage_female',
- 'unemployment_male',
- 'unemployment_female',
- lived in different address 1 year ago (male)
- lived in different address 1 year ago (female)
- lived in different address 5 years ago (male)
- lived in different address 5 years ago (female)

Historical Flood Records:

The historical flood records inventory consisted of 610 instances of floods that occurred in Queensland dating back to 2000, provided by the Australian Bureau of Meteorology's open data portal.[8,9] The raw data was provided as a table of text descriptions, each description containing flooded locations and their severities. These text tables of flood records were manually processed to obtain the final list of 610 distinctive flooding events, each corresponding with a time, location (both postal code and coordinates), and severity (minor, moderate, or major). It is worth noting that existing professional flood hazard modeling researchers consider a flood inventory size of 100-300 sufficient to analyze for trends.[10,11] Locations of non-flooding were selected using the precise buffer extent function in ArcGIS, which drew a boundary around each flooded location. We then randomly selected points outside of this boundary as locations which most likely did not experience flooding since 2000. The random selection estimation is necessary as there does not exist databases which record instances of non-flooding.

Hydrological Flooding Factors:

An analysis of past literature on conducting flood susceptibility modeling using neural networks was conducted to determine significant flood and hydrological factors to gather and input into the DLNN.[5,12,13] The 11 hydrological factors chosen were Elevation, Slope, Annual Rainfall, Annual Temperature, 9am humidity, 3pm humidity, Distance to river, Stream power index (SPI), Sediment transport index (STI), Topographic wetness index (TWI), Terrain ruggedness index (TRI). The hydrological factors were then processed using ArcGIS as raster datasets and the specific raster values at flooding locations and non-flooding locations were extracted. The benefit to using a GIS model is the ability to integrate various kinds of data.[14] The end result is a 1200+ entry dataset containing flooded locations, non-flooded locations, and hydrological factors at each location. This combined dataset with be referred to as the floods & factors dataset (see Figure 2 below) and is used to predict floods later in the paper.

No.	Flooded	Latitude	Longitude	Elevation	Slope	Aspect	Annual Rainfall
0	1	1	-26.405089	146.240425	297	0.090246	159.5672
1	1	1	-28.073536	145.687471	191	0.020194	189.4623
2	2	1	-28.848112	144.477449	132	0.034751	225
3	2	1	-27.760868	144.025295	140	0.030085	26.56505
4	2	1	-27.997093	143.819599	133	0.009116	281.3099

No.	Annual Mean Temp.	9am Humidity	3pm Humidity	Distance to River	SPI	STI	TWI	TRI
0	478.27	20.83917	50.957	29.904	0	0.000307	0.412745	5.001649
1	391.99	21.13125	51.159	30.681	0	0	0	4.70706
2	321.55	21.18958	50.012	30.516	0	0.001775	0.606998	8.494961
3	306.29	21.81583	47.205	29.222	0	0.021802	2.471373	11.27915
4	307.36	21.83792	47.029	29.373	0	0	0	5.502446

Figure 2: Floods & factors dataset, dataframe header displayed using python's pandas library.

The variables elevation, slope, distance to river, SPI, STI, TWI and TRI are constant. The variables annual rainfall, annual temperature and 9am/3pm humidity are annual climatological averages of the past 30 years. As such, this set of

flooding factors can be used to predict all flood entries within the floods & factors dataset.

Australian Census Data:

The Australian Bureau of Statistics publishes census data on their open data portal in five-year intervals, starting from 1996.[15] This research focuses on census data from 2006, 2011, 2016 and 2021. Census data from each postal code in Queensland, in each census year, was downloaded in the form of a detailed 50+ sheet excel summary of every census statistic in each postal code. Across four census years this totaled out to be 1800 such summaries. From each summary, the specific data entries that were extracted were: Mean age, Family income, Mortgage repayment, Rent, Population*, Indigenous percentage*, Unemployment rate*, Number of people who lived in a different address 1 year/5 years ago*, Place of usual residence 1 year/5 years ago*, and Industry of employment*. An asterisk (*) indicates the data is split into male & female counterparts. Specifically, census data on place of usual residence in the past will be critical in our attempt to understand human movement patterns as caused by floods.[16] During this data extraction process it was identified that some census files are either corrupt or have missing entries. The final inventory of census data used to infer the socio-economic impact of floods contains 1711 rows, spanning all postal codes of Queensland, Australia across four census years, each row complete with all the aforementioned census entries.

Methods

Flood Prediction – Unsupervised Methods:

To conduct preliminary analysis on the floods & factors dataset, this research ran agglomerative clustering and k-means clustering algorithms on the floods & factors dataset, to see if flooded/non-flooded locations share statistical similarities. This was done using the python scikit-learn library.[17] Each clustering algorithm was run once to cluster the entire dataset and once again to cluster highest contributing elements as determined by a principal component analysis (PCA). This research chose to use 5 clusters, with the rationale being these 5 clusters will hopefully represent: **no flood, minor flood, moderate flood, major flood, and a miscellaneous unforeseen category.**

Agglomerative clustering is a hierarchical clustering algorithm that starts with each data point as its own individual cluster.[18] The algorithm then iteratively computes distances between sets, and the closet clusters are merged. This process is repeated until only a set number of clusters remain.

K-means clustering begins by picking a random selection of K vectors (K datapoints, as each datapoint can be thought of as an N-dimensional vector) to be centroid vectors. Then the clustering algorithm determines the closest centroid for each datapoint, a process that splits the data into K groups.[19] The centroid of every group is then recalculated and each datapoint is again grouped to the closest centroid. This process iterates until the dataset eventually converges on K clusters. By using agglomerative clustering which is a hierarchical method and K-means clustering which is a partition-based method, this research can circumvent K-means clustering's tendency to draw to local optima, and agglomerative clustering's high time complexity.[20]

A **principal component analysis** works by transforming the entire existing dataset into an N-dimensional vector space. This vector space is then transformed, using matrix multiplication, into a new vector space with the same dimensionality. At this point, each of the N new vectors in the new vector space is a linear addition of contributions from the N original vectors. Contribution here refers to how much each variable affects the variance in the data. We then obtain a list of contributions and choose the highest ones until we reach some threshold (e.g. when a combination of contributions add up to >95%, we can discard the rest). Clustering results are in the format (% flooded areas / % unflooded areas):

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Agglo.	0/100	100/0	66/33	100/0	95/5
K-means	85/15	100/0	0/100	0/100	100/0
Agglo. w/PCA	0/100	55/45	100/0	100/0	95/5
K-means w/PCA	100/0	0/100	5/95	0/100	100/0

Figure 3: Clustering results of floods & factors dataset.

All four unsupervised clustering methods have grouped data into some clusters that contain solely flooded or non-flooded areas. This is evidence that there exists some distinction that separates flooded and non-flooded locations, enabling our dataset to be analysed using supervised methods.

Flood Prediction – Supervised Methods:

The 1200+ entries within the floods & factors dataset were assigned a binary classifier of 0 and 1, representing no flooding and flooding respectively. A correlation heatmap was then generated using the seaborn python library. From this heatmap we can infer that the flooding factors considered in this research are sufficiently distinctive from each other and are valid to use as unique predictors, as most correlation coefficients are between -0.5 to +0.5, representing low to negligible correlation.

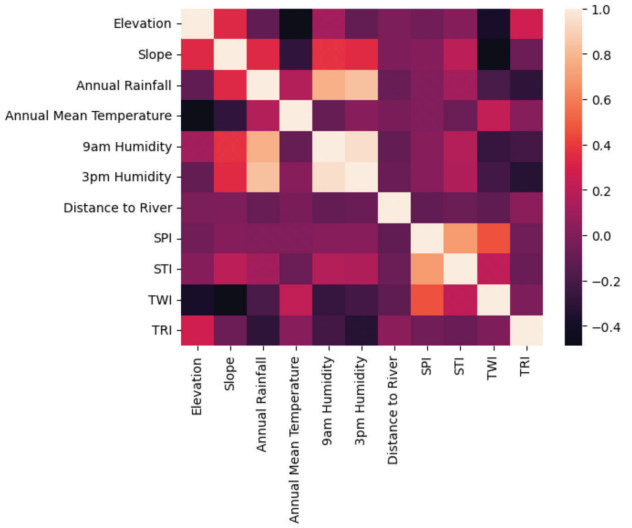

Figure 4: Correlation heatmap of hydrological flooding factors.

DOI: 10.36838/v6i12.4

The floods & factors dataset was then passed through scikit-learn's StandardScaler,[21] which is used for rescaling the factors so that they have the properties of a standard normal distribution with a mean of 0 and a standard deviation of 1. This is also known as z-score normalization. The scaled dataset is then split into 70% training data and 30% testing data.

	Elevation	Slope	Annual Rainfall	Annual Mean Temperature	9am Humidity	3pm Humidity	Distance to River	SPI	STI	TWI	TRI
0	0.509300	-0.369572	-0.388525	-0.850584	-0.588973	-0.651970	-0.707113	-0.213074	-0.188460	0.385241	0.105521
1	-0.163554	-0.543991	-0.577033	-0.728614	-0.570088	-0.584511	-0.707113	-0.281104	-0.407978	0.253786	-1.661156
2	-0.538067	-0.507746	-0.730933	-0.704256	-0.677320	-0.598836	-0.707113	0.112229	-0.085147	1.944065	1.855657
3	-0.487286	-0.519364	-0.764274	-0.442741	-0.939745	-0.711180	-0.707113	4.550140	0.906418	3.186456	-0.179347
4	-0.531720	-0.571574	-0.761936	-0.433517	-0.956199	-0.698071	-0.707113	-0.281104	-0.407978	0.608712	-0.450194

Figure 5: Header of scaled dataset using sklearn's StandardScaler.

Flood Prediction – K Nearest Neighbours Model:

The K Nearest Neighbour (KNN) algorithm works by finding the decision boundary that separates one class from another (i.e. flooded vs non-flooded). Each entry in the floods & factors dataset can be represented as an 11-dimensional vector. Given a certain vector/entry, the KNN model determines the distance of X to every vector in the dataset. Then, the K vectors that are closest (nearest neighbours) to X are selected, and the class that the most neighbours belong to is the predicted class of X. When given the full set of 11 flooding factors, the KNN model is roughly 80% accurate in predicting whether a flood has occurred based on local hydrological variables. However, this may not be the optimal set of variables to predict for floods. To solve this, we took the novel approach of iteratively running the KNN model with different combinations of 2 to 11 flood factors, and picking the top performing sets of combinations. This is done using the combinations() function of python's itertools module.

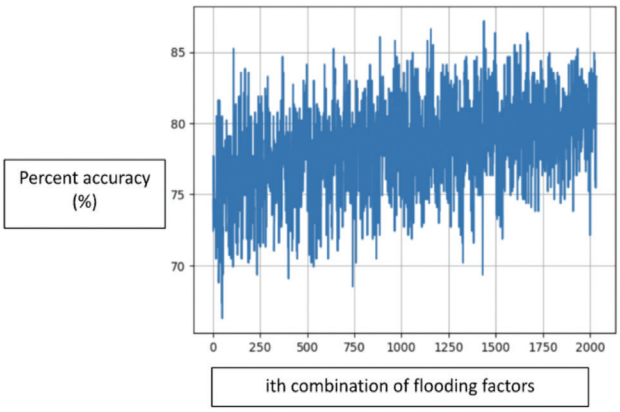

Figure 6: Graph of KNN model accuracies for all combinations of flooding factors.

As seen in Figure 6, the KNN model performs to a maximum of just below 90% accuracy, between the 1250th and 1500th combination, when using an optimal set of variables of **'Annual Rainfall'**, **'9am Humidity'**, **'Distance to River'**, **'STI'**, **'TWI'**, and **'TRI'**. This accuracy serves to roughly estimate the accuracy by which the neural network can draw predictions from this dataset and will be later verified by the neural network.

Flood Prediction – Deep Learning Neural Network:

The decision to use a Deep Learning Neural Network was inspired by Mosavi *et al.*'s study into major machine learning methods used in present flood prediction literature.[22] Mosavi identified that usage of Artificial Neural Networks, and by extension Deep Learning Neural Networks, far exceed that of other more interpretable ML methods such as Support Vector Machines, Decision Trees, and Random Forest. A **Neural Network** works by establishing an input layer, an output layer, and hidden layers of neurons. In this research, as each neuron is connected to each pervious neuron, the network is "dense". Each connection between neurons is a multiplication by a weight, and for each layer, each neuron is added to a bias. At the output layer an activation function - for this research the Rectified Linear Unit or ReLU function was used - outputs a value between 0 and 1.[23] Then, the loss, or the difference between the network's performance and the actual result, is computed as a function of the network weights. To minimize loss, we can determine its gradient, which is a partial derivative of the loss function with respect to weight. We then shift our weights in the opposite direction of the gradient in a process called gradient descent. By iteratively running the neural network over a set number of epochs (100 epochs for this research), we can loop until the loss converges at a minimum, by which point the neural network is trained.

```
#build the NN model ...
nn_model = Sequential()

n_neurons1 = 50
n_neurons2 = 50
nn_model.add(Dense(n_neurons1, input_dim = X_train.shape[1], activation= 'relu'))  #vary # of neurons
nn_model.add(Dense(n_neurons2, activation= 'relu'))   #vary # of neurons
nn_model.add(Dense(y_cat.shape[1], activation = 'softmax'))  #Softmax means output probability of each possible digi

#compile the model
nn_model.compile(loss='categorical_crossentropy', metrics=['accuracy'], optimizer='adam')
#Loss function: Cross entropy is used with probabilities suchas softmax
#Accuracy = Number of correct predictions/total number of predictions

#train the model ...
nn_model.fit(X_train, y_cat[train_indices,:], epochs = 100, batch_size=50)
```

Figure 7: Deep learning neural network architecture.

The deep learning neural network used in this research consisted of an input layer of 11 flooding factors, two hidden layers each with 50 neurons, and one output layer that is an estimation of either 1 or 0. 1 representing a flood prediction, 0 representing a no-flood prediction. A dense network architecture is used in which every neuron is connected to every neuron in the previous layer, and the Rectified Linear Unit (ReLU) activation function is used to introduce non-linearity to the model. The rationale of only using two hidden layers was to prevent overfitting the DLNN, as a two-hidden-layer DLNN already reaches training accuracies of upwards of 98%. This DLNN architecture is similar to that of Kalantar *et al*'s 2021 study.[10]

```
#make predictions ...
nn_pred_raw = nn_model.predict(X_test)
nn_pred = np.argmax(nn_pred_raw, axis=1)   #find indexof the output neuron with the maximum value
print(np.sum(nn_pred == df2['cat_Flooded'].iloc[test_indices].to_numpy())/len(nn_pred))
```

Figure 8: Deep learning neural network accuracy.

Here we let the trained DLNN make predictions on X_test, which is the testing set comprised of 30% of the original dataset. The second line "np.argmax(nn_pred_raw, axis=1)" identifies the index of the maximum value in each row, corresponding to the predicted class label for that sample. Since this DLNN is a binary classifier, the class of 0 represents no flood and the class of 1 represents flood. These predictions are stored

in nn_pred, which is then printed. Multiple executions of the DLNN return a prediction accuracy of around 85-90%, which verifies the prior optimised KNN model's result. It is worth noting that existing flood hazard researchers generally accept 80-90% model accuracy.[1,13,24] The highest model accuracy that we have seen is in Kalantar *et al*'s research – 92.11% – albeit obtained with more complex methods, such as particle swarm optimisation and long-short term memory models.[10]

Socioeconomic Inference – Introduction:

Current research on flood hazard modelling focuses heavily on data analysis, implementing predictive models, and fine-tuning parameters within these models.[4,10-12,24,25] All the above papers end their research section when their respective researchers have successfully implemented a model that is accurate to an acceptable standard. However, significantly less focus is placed on the socio-economic impacts of floods once they are predicted to occur. This research identifies the current lack of quantitative research conducted on the socio-economic impacts of flooding events, and we seek to innovate in the field of flood hazard assessment by bridging interdisciplinary fields of environmental science and sociology, the benefits of this interdisciplinary approach in Environmental Science being already recognized by many researchers such as Pacifico, A.[26] To our knowledge, this research is the first ever attempt to analyse Australian flood data and census data simultaneously. Using an ordinary least squares (OLS) regression model, we have determined how floods impact sociological variables such as rent, income, unemployment, Indigenous populations, and more. We first combined the historical flood records dataset with the Australian Census Data dataset (see subsections respectively under **Data Acquisition and Processing**) via data merging. This research rejects the null hypothesis at an alpha value of 0.05.

Socioeconomic Inference – Data Merging:

The two datasets used to infer the socio-economic impacts of floods are the Australian census dataset and the historical flood records dataset. They are left merged with Australian census data on the left. A left merge is a way of joining two datasets that keeps all data from the left dataset. Rows from the right dataset are combined with rows from the left dataset based on a common column. In this research, the common column is zipcode, which both represents flood locations and census locations.

- For every row in census data, if census zipcode is the same as a zipcode in flood data, data from that flood data row will be joined to the right of the census data row.

- If a census zipcode matches multiple flood data zipcodes (which it often does), duplicates of the census row are created, each one matching a different flood row.

- If a census zipcode does not have a matching flooded data zipcode (i.e. the census area has not been flooded over the last 20 years), a row of NaN (Not a Number) values are placed to the right of the census row.

Socioeconomic Inference – Categorical Dummy Variables:

Before we can run the Ordinary Least Squares (OLS) regression model on this merged dataset, we need some way to determine whether a zipcode has experienced a flood prior to a census year, when the census data was taken. We appended a final boolean column "C(flooded_before_census)" to the merged dataset. The "C" denotes that a variable is categorical.

- **C(flooded_before_census) = 1** means a flood has occurred in a zipcode, and a census was taken afterwards. The census data would therefore reflect impacts attributable to the flood.

- **C(flooded_before_census) = 0** means one of two things: either the census was taken before the flood had occurred, or the census was taken in an area where no floods had occurred. In either case, this data entry will not reflect impacts attributable to the flood.

By using C(flooded_before_census) as a dummy explanatory variable in an OLS regression line, we can look at its coefficient to analyse the impact of floods on the selected output variable.[27] The coefficient of C(flooded_before_census) represents how big is the difference in a selected census variable (e.g. rent) between flooded and non-flooded areas, thus giving us the societal impact of floods. Given the extensive size of our merged dataset (3562 rows of census data), we are also able to interpret this impact at a high confidence level.

Aside from C(flooded_before_census), this research will use two additional categorical variables in the OLS regression analysis: C(census_year) is a categorical variable denoting the year that a census was taken, and C(zipcode) is a categorical variable denoting the zipcode where a census was taken. The reason for using C(census_year) and C(zipcode) is to control for, respectively, underlying time & space fixed effects which may influence the trend demonstrated by our OLS regression lines.

Time fixed effects are external effects that change over time but are constant across every point in space at any time, examples are government policy and global climate patterns. Space fixed effects are external effects that vary across space but are constant throughout time at any given space, examples are neighborhood socio-economic differences, distance to major cities, and local geography.[28]

Let's say we wish to estimate the impact of flooding on a census statistic like rent. When obtaining a trend from OLS regression, we need to be cautious whether any predicted increase/decrease in rent was actually caused by flooding, or by external factors such as fiscal economic policy (a time fixed effect) and/or local housing price differences (a space fixed effect).[29] It is possible that the rent decline was a normal part of the markets and not due to flooding.

To control for the effects of these underlying trends, we must add C(census_year) and C(zipcode) as dummy variables in the regression line, which allows the OLS regression model to assign coefficients to the specific year and zipcode of each datapoint that the regression line uses. The coefficient of C(census_year) will represent how time-variable external effects will have affected the regression each year, and the coefficient of C(zipcode) will represent how space-variable external effects will have affected the regression each zipcode. Now if

DOI: 10.36838/v6i12.4

we examine the coefficient of C(flooded_before_census), we will have reached a more unbiased estimate of the societal impact of floods, where we have controlled time & space fixed effects to the best of our ability.

Socioeconomic Inference – Ordinary Least Squares Regression:

An Ordinary Least Squares (OLS) regression line attempts to establish a linear relationship between the explanatory variables (X_i) and the output variable (Y) such that:

$$Y = a_o + a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_nX_n + error$$

Where a_0 is the intercept of the regression line, and the error term attempts to capture unobserved factors. The goal is to minimize the sum of squared differences between the regression's output and the actual dataset value. We can work out the regression coefficients (a_i) as such:

$$a_i = \sum \left[(X_i - X_{i\mu})(Y - Y_\mu)\right] \div \sum \left(X_i - X_{i\mu}\right)^2$$

Where $X_{i\mu}$ represents the mean of the i^{th} explanatory variable, and Y_μ represents the mean of the output variable. OLS regression was performed using the Statsmodels python package. The scaffold code for each regression line is as follows:

```
reg_model = smf.ols(
    formula='y_variable ~ x_variable + x_variable + ... + C(Zipcode) + C(census_year) + C(flooded_before_census)',
    data=sample).fit()

print(reg_model.summary())
```

Figure 9: OLS regression baseline code.

When running a regression line, y_variable is replaced with the chosen census variable to predict. The multiple instances of x_variable are each replaced with a different census variable that is correlated with the y_variable. reg_model.sumary() returns a table of coefficients for each x_variable. The coefficient of the dummy variable C(flooded_before_census) is then interpreted as the impact of flooding on the y_variable. OLS Regression results are as follows:

	y_variable (response variable)	Coeff. of C(flooded_before_census)	Adj. R^2	P value	F stat.	ANOVA
A	Median household income ($AUD weekly)	-35.4763	0.890	0.000	64.98	All Three
B	Median rent ($AUD weekly)	-6.4744	0.923	0.002	96.28	All Three
C	Median housing loan repayment ($AUD monthly)	-38.5837	0.615	0.099	13.59	All Three
D	Average household size	0.8042	0.146	0.132	2.354	All Three
E	% Unemployment (female)	0.0021	0.655	0.139	15.96	All Three
F	% Unemployment (male)	0.0015	0.663	0.312	16.55	All Three
G	Population (female)	-39.4133	0.981	0.537	416.9	All Three
H	Population (male)	-55.6891	0.983	0.351	455.4	All Three
I	Indigenous count (female)	13.8887	0.902	0.126	74.05	All Three
J	Indigenous count (male)	15.6013	0.902	0.080	73.78	All Three
K	Indigenous percentage (female)	0.0056	0.772	0.263	27.69	Zipcode
L	Indigenous percentage (male)	0.0059	0.826	0.127	38.57	Zipcode, Census Year
M	# of people who lived in a different address 1 year ago (female)	-17.8885	0.974	0.185	301.0	All Three
N	# of people who lived in a different address 1 year ago (male)	-32.3528	0.974	0.011	300.7	All Three
O	# of people who lived in a different address 5 years ago (female)	-31.4886	0.973	0.340	288.1	All Three
P	# of people who lived in a different address 5 years ago (female)	-55.5624	0.975	0.065	303.8	All Three

Figure 10 A-P: Results of regression lines with C(flooded_before_census), C(census_year) and C(Zipcode) as explanatory variables.

	y_variable (response variable)	Coeff. of C(flooded_before_census)	Adj. R^2	P value	F stat.	ANOVA
A	Median household income ($AUD weekly)	283.0209	0.624	0.000	14.18	Both
B	Median rent ($AUD weekly)	64.4725	0.795	0.000	31.77	Both
C	Median housing loan repayment ($AUD monthly)	164.0531	0.569	0.000	11.47	Both
D	Average household size	1.1973	0.137	0.009	2.260	Both
E	% Unemployment (female)	0.0066	0.609	0.000	13.38	Both
F	% Unemployment (male)	0.0103	0.617	0.000	13.83	Both
G	Population (female)	540.1989	0.979	0.000	369.8	Both
H	Population (male)	484.4042	0.981	0.000	406.3	Both
I	Indigenous count (female)	65.7023	0.896	0.000	69.57	Both
J	Indigenous count (male)	67.7436	0.895	0.000	68.93	Both
K	Indigenous percentage (female)	0.0070	0.772	0.097	27.87	Zipcode
L	Indigenous percentage (male)	0.0130	0.825	0.000	38.51	Both
M	# of people who lived in a different address 1 year ago (female)	21.1615	0.974	0.067	295.8	Zipcode
N	# of people who lived in a different address 1 year ago (male)	-9.6255	0.974	0.376	298.9	Zipcode
O	# of people who lived in a different address 5 years ago (female)	117.8125	0.972	0.000	273.4	Both
P	# of people who lived in a different address 5 years ago (female)	61.9602	0.973	0.017	292.8	Both

Figure 11 A-P: Regression analysis with only C(flooded_before_census) and C(Zipcode) as explanatory variables.

Regression lines highlighted in green are highly significant and have p value < 0.05, adjusted R squared > 0.80.

Regression lines highlighted in orange are moderately significant and have p value < 0.1, adjusted R squared > 0.60.

Regression lines that are not highlighted as deemed not statistically significant as are not considered for interpretation.

This research also ran Two-Way and Multi-Way ANOVA to determine whether the aforementioned categorical variables are statistically significant. The ANOVA column describes the respective categorical variables in Figures 10 & 11 respectively that **have a significant impact on the y_variable**. A sample of the ANOVA code is shown below:

```
# Specify the formula for Two-Way ANOVA with interaction
formula = 'median_rent ~ C(flooded_before_census) + C(Zipcode) + C(census_year)'

# Fit the model
model = ols(formula, data=sample).fit()

# Perform ANOVA
anova_table = sm.stats.anova_lm(model, typ=2)  # Type 2 ANOVA for balanced designs
print(anova_table)
```

```
                          sum_sq      df           F    PR(>F)
C(flooded_before_census)  9.970535e+03     1.0    10.088665  0.001506
C(Zipcode)                3.062237e+07   447.0    69.318084  0.000000
C(census_year)            5.180621e+06     3.0  1747.333587  0.000000
Residual                  3.073584e+06  3110.0          NaN       NaN
```

Figure 12: Sample Three-Way ANOVA code snippet.

While other researchers may decide otherwise, this research rejects the null hypothesis at an alpha value of 0.05.[30] A regression line on any given y_variable was only considered significant if the p value of said regression line met this alpha value (p < 0.05). A common trend identified as we undertook regression analysis is that by including all three categorical variables of flooding before census, census year and census zipcode, the resultant regression line often contains a p value greater than 0.05. If we were to conduct regression analysis with only categorical variables for flooding before census and census zipcode, without using census year, we can significantly reduce the p values for regression lines that previously had p > 0.05, such as "Population", "Indigenous Count", and "Number of people who lived in a different address 5 years ago".

However, regression lines for "Median household income" and "Median rent" saw a significant decrease in adjusted R squared when census year was removed as an explanatory variable. The implication for this result is that "Median household

income" and "Median rent" are affected by underlying trends across time that need to be controlled by a categorical variable - C(census_year) - to yield statistically significant regression lines with R^2 > 0.85 and p < 0.05. While census statistics such as "Population" and "Indigenous Count" are less affected by such trends, hence C(census_year) could not be used as an effective explanatory variable for these statistics.

Result and Discussion

Interpreting Regression Coefficients:

The OLS regression model's results (Figures 10 & 11) can be interpreted as such:

For **highly significant** regression lines (highlighted in **green** in Figures 10 & 11):

- Households that have experienced floods earn **$35.4763 dollars AUD less** per week. (Figure 10A)
- Households that have experienced floods pay **$6.4744 dollars AUD less** in rent per week. (Figure 10B)
- The number of males who lived in a different address 1 year ago **decreases by 32.3528** in each flooded zipcode. (Figure 10N)
- The female population **increases by 540.1989** in each flooded zipcode. (Figure 11G)
- The male population **increases by 484.4042** in each flooded zipcode. (Figure 11H)
- The Indigenous female population **increases by 65.7023** in each flooded zipcode (Figure 11I)
- The Indigenous male population **increases by 67.7436** in each flooded zipcode (Figure 11J)
- The Indigenous male percentage **increases by 0.0130** in each flooded zipcode (Figure 11L)
- The number of females who lived in a different address 5 years ago **increases by 117.8125** in each flooded zipcode. (Figure 11O)
- The number of males who lived in a different address 5 years ago **increases by 61.9602** in each flooded zipcode. (Figure 11P)

For **moderately significant** regression lines (highlighted in **orange** in Figures 10 & 11):

- Household that have experienced floods pay **$38.5837 dollars AUD less** in mortgage repayment per month (Figure 10C)
- Female unemployment percentage **increases by 0.0066** in each flooded zipcode (Figure 11E)
- Male unemployment percentage **increases by 0.0103** in each flooded zipcode (Figure 11F)
- The number of females who lived in a different address 1 year ago **increases by 21.1615** in each flooded zipcode. (Figure 11M)

In the event that both a highly significant regression line and a moderately significant regression line existed for the same census variable (y_variable), only the highly significant regression line is interpreted.

Evidence for Environmental Redlining in Australia:

We have highly statistically significant results of how flooding affects the total population (by sex) and the Indigenous population (by sex) at the zipcode level. We can further conduct the following analysis:

Dividing the female Indigenous population increase (65.7023) by the total female population increase (540.1989) gives the result that Indigenous females account for (65.7023/540.1989)*100 = **12.16%** of the total female population increase in a flooded zipcode. Similarly, dividing the male Indigenous population increase (67.7436) by the total male population increase (484.4042) gives the result that Indigenous males account for (67.7436/484.4042)*100 = **13.98%** of the total male population increase in a flooded zipcode. As of 2021 when the latest Australian census was conducted, Indigenous Australians only account for 3.8% of Queensland's total population.[31] Assuming that floods impact Indigenous Australians and non-Indigenous Australians in the same way, we should expect the percentage increase in Indigenous population to mirror the 5.2% of Indigenous Australians in Queensland. However, the actual increase in Indigenous population does not reflect this.

Indigenous Australians only make up 3.8% of the Australian population but account for 13% of the population increase in areas that have experienced flooding. This is evidence that flooding disproportionately impacts Indigenous Australians.

Limitations and Considerations - Unsupervised Methods:

Even though agglomerative clustering and k-means clustering are mathematically different,[20] both were able to ascertain what appears to be a decision boundary between flooded and non-flooded areas, as demonstrated by Figure 3. Not only does this explain the success of our deep learning neural network as a binary classification model, but it also demonstrates that our historical flood dataset is statistically valid to be analysed by other, less interpretable models, such as K-Nearest Neighbour predictor or a Neural Network.[32]

Limitations and Considerations - Supervised Methods:

A similar statement can be made here – that K nearest neighbor models and neural networks are mathematically different, yet both models returned an accuracy of 85-90%, reinforcing the reliability of our predictions. Given any geographical coordinate in Queensland and its corresponding set of hydrological flooding factors, we can determine whether a flood likely occurred in said location. This result not only applies to the past 20 years within the floods dataset, it is also extrapolatable into the future. This research does not use time-series data to make predictions. For data elements such as rainfall, temperature, and humidity that does change over time, this research takes the 30-year climatological average (1990-2020) provided by the Bureau of Meteorology.[9] In essence, the neural network sees climate as a static, geographically-dependent factor, and thus a prediction of flooding in an area can be generalised to imply "the flooding factors of this area is such that a flood is likely to occur, or to have occurred". A limitation of this research is that climatological variables are obviously not static, and the neural network's predictions will not be forever accurate, but this is to be expected of most prediction models that rely on past mete-

DOI: 10.36838/v6i12.4

orological data.[33] Should the world's climate change drastically in the future, different datasets of more recent climatological averages will need to be used to update the model.

Limitations and Considerations – Regression Model:

When running the OLS Regression Model, we minimised selection bias to the best of our ability by keeping the state level (Queensland) constant across all census data and flood data observations, and by collecting census data from every postal code in Queensland. Further, this research has accounted for **time and space fixed effects** through the inclusion of census year & zipcode, respectively, as dummy variables in the regression line (see subsection under **Methods – Socioeconomical Inference**). One significant limitation of OLS regression is the multicollinearity of input variables.[34] When multiple explanatory variables are strongly correlated with each other and hence difficult to distinguish apart, it is difficult to determine the regression coefficients of each. We can prevent this by choosing explanatory variables with low correlations.[35]

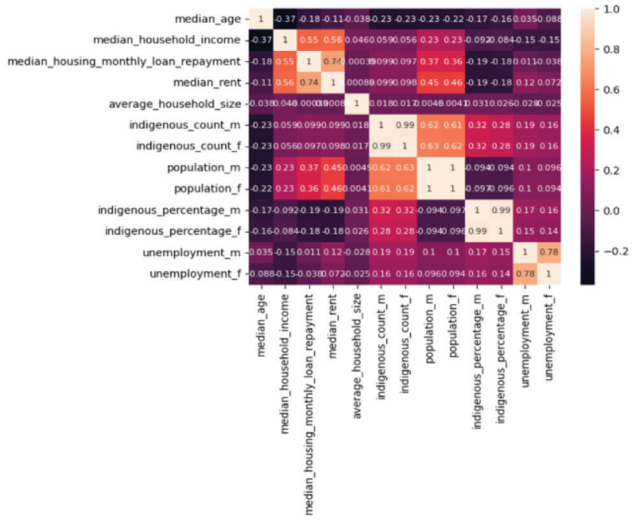

Figure 13: Correlation matrix including all explanatory Australian census data variables.

While median_rent and median_housing_monthly_loan_repayment exhibit moderate to high correlation, the majority of explanatory variables used in OLS regression have low correlation with each other. The correlation of the same variable between different genders is irrelevant, as no regression line in Figures 10 & 11 uses both male and female data. Another limitation of OLS regression is that it can only be applied to explanatory variables that have a linear correlation with the output variable, meaning that OLS regression cannot assess the impact of floods on all census data variables. An example non-linear variable is median_age, for which no regression line was run.

Future Research Directions:

While this research attempts to ascertain the impacts of flooding on census statistics, the OLS regression model we employed assumes - for the sake of simplicity - that all flood events are identical. Future research could make use of the "Level" classification in the Historical Flood Records dataset, which separates floods into "minor", "moderate", and "major" in terms of severity defined by the Australian Government.[36] This could be done by considering flood level as a categorical dummy variable in the OLS regression line.

Alternatively, future research could use a similar DLNN architecture to train a more nuanced flood prediction model that can determine the severity of an upcoming flood. It should be noted that the researched attempted this process yet was only able to achieve a DLNN accuracy around 80% due to a lack of data. Should future research be undertaken in this direction, it is recommended to scout for additional historical flood records outside of Queensland, Australia, preferably from other Australian States along the eastern coast. Furthermore, this research only decided to use 11 hydrological flooding factors as predictor variables, as these variables were introduced in Kalantar *et al.*'s research and preduced effective results with model accuracies up to 92%.[10] Additional climatological and/or geological variables, such as ENSO climate patterns, soil moisture, soil permeability etc were not considered in the scope of this research, and can become great refinements to this flood predicting model in the future.

Regarding societal concerns and policymaking for the discovered disproportionate impacts of floods, the author has reached out to representatives from the Australian New South Wales Government Department of Housing, Planning and Infrastructure, as well as the Indigenous Australian social enterprise Currie Country Social Change in a discussion of this matter. While it isn't realistic to move everyone to live away from flood prone areas, this research - alongside many flood reports undertaken by the Australian government will nevertheless serve as a reminders for the ignorance of Indigenous knowledge of flood areas during colonel Australian land development. Hence this research seeks to empower a strive in policymaking towards embracing traditional wisdom for future infrastructure development.

■ Conclusion

In this research, we compiled the largest historical flood dataset in Australian flood research, mapping flood locations over two decades using ArcGIS software and creating raster datasets of hydrological factors. Unsupervised clustering algorithms testament to the statistical significance of our dataset. And an 85-90% DLNN accuracy demonstrates our success at predicting floods. Additionally, combining interdisciplinary approaches of environmental science and sociology, this research used ordinary least squares regression to reveal potential evidence of environmental redlining towards Indigenous Australians. The regression model used by this research can also quantify the socio-economic impact of floods in Queensland. Through this research, we hope to lay a stepping stone on humanity's path towards flood prevention, which future researchers will hopefully find useful.

■ Acknowledgments

This research would like to acknowledge Morgan State University and Duke University for their continued support

throughout the research process. This research would also like to thank Currie Country Social Change for informing the Australian Parliament of this research and similar studies on the impacts of natural disasters on Indigenous Australians. Additionally, we would like to thank the International Science & Engineering Fair (ISEF) and the Chinese Adolescents Science and Technology Innovation Contest (CASTIC) for their recognition of this project with a Third Place Grand Award and First Place Award respectively.

References

1. Kumbier, K.; Carvalho, R. C.; Vafeidis, A. T.; Woodroffe, C. D. Investigating Compound Flooding in an Estuary Using Hydrodynamic Modelling: A Case Study from the Shoalhaven River, Australia. Natural Hazards and Earth System Sciences 2018, 18 (2), 463–477. https://doi.org/10.5194/nhess-18-463-2018
2. Johnson, F.; Hutchinson, M. F.; The, C.; Beesley, C.; Green, J. Topographic Relationships for Design Rainfalls over Australia. Journal of Hydrology 2016, 533, 439–451. https://doi.org/10.1016/j.jhydrol.2015.12.035
3. ENGELS, B. Capital Flows, Redlining and Gentrification: The Pattern of Mortgage Lending and Social Change in Glebe, Sydney, 1960-1984. International Journal of Urban and Regional Research 1994, 18 (4), 628–657. https://doi.org/10.1111/j.1468-2427.1994.tb00290.x
4. Kelly, M.; Schwarz, I.; Ziegelaar, M.; Watkins, A. B.; Kuleshov, Y. Flood Risk Assessment and Mapping: A Case Study from Australia's Hawkesbury-Nepean Catchment. Hydrology 2023, 10 (2), 26. https://doi.org/10.3390/hydrology10020026
5. Kourgialas, N. N.; Karatzas, G. P. Flood Management and a GIS Modelling Method to Assess Flood-Hazard Areas—a Case Study. Hydrological Sciences Journal 2011, 56 (2), 212–225. https://doi.org/10.1080/02626667.2011.555836
6. Yousefi, S.; Pourghasemi, H. R.; Emami, S. N.; Pouyan, S.; Eskandari, S.; Tiefenbacher, J. P. A Machine Learning Framework for Multi-Hazards Modeling and Mapping in a Mountainous Area. Scientific Reports 2020, 10 (1). https://doi.org/10.1038/s41598-020-69233-2
7. Antwi-Agyakwa, K. T.; Afenyo, M. K.; Angnuureng, D. B. Know to Predict, Forecast to Warn: A Review of Flood Risk Prediction Tools. Water 2023, 15 (3), 427. https://doi.org/10.3390/w15030427
8. Australian Bureau of Meteorology. Queensland Flood History. www.bom.gov.au. http://www.bom.gov.au/qld/flood/fld_history/index.shtml
9. Australian Bureau of Meteorology. Maps of average conditions. www.bom.gov.au. http://www.bom.gov.au/climate/maps/averages/
10. Kalantar, B.; Ueda, N.; Saeidi, V.; Janizadeh, S.; Shabani, F.; Ahmadi, K.; Shabani, F. Deep Neural Network Utilizing Remote Sensing Datasets for Flood Hazard Susceptibility Mapping in Brisbane, Australia. Remote Sensing 2021, 13 (13), 2638. https://doi.org/10.3390/rs13132638
11. Lei, X.; Chen, W.; Panahi, M.; Falah, F.; Rahmati, O.; Uuemaa, E.; Kalantari, Z.; Ferreira, C. S. S.; Rezaie, F.; Tiefenbacher, J. P.; Lee, S.; Bian, H. Urban Flood Modeling Using Deep-Learning Approaches in Seoul, South Korea. Journal of Hydrology 2021, 601, 126684. https://doi.org/10.1016/j.jhydrol.2021.126684
12. Gude, V.; Corns, S.; Long, S. Flood Prediction and Uncertainty Estimation Using Deep Learning. Water 2020, 12 (3), 884. https://doi.org/10.3390/w12030884
13. Mia, Md. U.; Rahman, M.; Elbeltagi, A.; Abdullah-Al-Mahbub, Md.; Sharma, G.; Islam, H. M. T.; Pal, S. C.; Costache, R.; Islam, A. R. Md. T.; Islam, M. M.; Chen, N.; Alam, E.; Washakh, R. M. A. Sustainable Flood Risk Assessment Using Deep Learning-Based Algorithms with a Blockchain Technology. Geocarto International 2022, 1–29. https://doi.org/10.1080/10106049.2022.2112982
14. Kumar, V.; Kul Vaibhav Sharma; Tommaso Caloiero; Mehta, D. J.; Singh, K. P. Comprehensive Overview of Flood Modeling Approaches: A Review of Recent Advances. Hydrology 2023, 10 (7), 141–141. https://doi.org/10.3390/hydrology10070141
15. Australian Bureau of Statistics. Search Census data - Australian Bureau of Statistics. www.abs.gov.au, 2021. https://www.abs.gov.au/census/find-census-data/search-by-area
16. Wesolowski, A.; Buckee, C. O.; Pindolia, D. K.; Eagle, N.; Smith, D. L.; Garcia, A. J.; Tatem, A. J. The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales. PLoS ONE 2013, 8 (1), e52971. https://doi.org/10.1371/journal.pone.0052971
17. Scikit-Learn. User guide: contents — scikit-learn 0.22.1 documentation. Scikit-learn.org. https://scikit-learn.org/stable/user_guide.html
18. Michener, C. D.; Sokal, R. R. A Quantitative Approach to a Problem in Classification. Evolution 1957, 11 (2), 130. https://doi.org/10.2307/2406046.
19. Phillips, J. University of Utah. K-Means Clustering. https://users.cs.utah.edu/~jeffp/teaching/cs5140-S14/cs5140/L10-kmeans.pdf
20. Yin, H.; Aryani, A.; Petrie, S.; Nambissan, A.; Astudillo, A.; Cao, S. A Rapid Review of Clustering Algorithms. arXiv.org. https://doi.org/10.48550/arXiv.2401.07389
21. Scikit-Learn. User guide: contents — scikit-learn 0.22.1 documentation. Scikit-learn.org. https://scikit-learn.org/stable/user_guide.html
22. Mosavi, A.; Ozturk, P.; Chau, K. Flood Prediction Using Machine Learning Models: Literature Review. Water 2018, 10 (11), 1536. https://doi.org/10.3390/w10111536
23. Nair, V.; Hinton, G. Rectified Linear Units Improve Restricted Boltzmann Machines; 2010. https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf.
24. Li, X.; Yan, D.; Wang, K.; Weng, B.; Qin, T.; Liu, S. Flood Risk Assessment of Global Watersheds Based on Multiple Machine Learning Models. Water 2019, 11 (8), 1654–1654. https://doi.org/10.3390/w11081654
25. Nkwunonwo, U. C.; Whitworth, M.; Baily, B. A Review of the Current Status of Flood Modelling for Urban Flood Risk Management in the Developing Countries. Scientific African 2020, 7, e00269. https://doi.org/10.1016/j.sciaf.2020.e00269
26. Pacifico, A. The Impact of Socioeconomic and Environmental Indicators on Economic Development: An Interdisciplinary Empirical Study. Journal of risk and financial management 2023, 16 (5), 265–265. https://doi.org/10.3390/jrfm16050265
27. Grace-Martin, K. Interpreting Regression Coefficients - The Analysis Factor. The Analysis Factor. https://www.theanalysisfactor.com/interpreting-regression-coefficients/
28. Schmelzer, C. H., Martin Arnold, Alexander Gerber, and Martin. 10.4 Regression with Time Fixed Effects | Introduction to Econometrics with R. https://www.econometrics-with-r.org/10.4-regression-with-time-fixed-effects.html
29. Firebaugh, G.; Warner, C.; Massoglia, M. Fixed Effects, Random Effects, and Hybrid Models for Causal Analysis. Handbooks of Sociology and Social Research 2013, 113–132. https://doi.org/10.1007/978-94-007-6094-3_7
30. Kwak, S. G. Are Only P-Values Less than 0.05 Significant? A P-Value Greater than 0.05 Is Also Significant! Journal of Lipid and Atherosclerosis 2023, 12 (2), 89–89. https://doi.org/10.12997/jla.2023.12.2.89

DOI: 10.36838/v6i12.4

31. Australian Institute of Health and Welfare. Profile of First Nations people. Australian Institute of Health and Welfare. https://www.aihw.gov.au/reports/australias-welfare/profile-of-indigenous-australians#:~:text=2018).-
32. AWS. Model interpretability - Machine Learning Best Practices in Healthcare and Life Sciences. docs.aws.amazon.com. https://docs.aws.amazon.com/whitepapers/latest/ml-best-practices-healthcare-life-sciences/model-interpretability.html
33. Alizadeh, O. Advances and Challenges in Climate Modeling. Climatic Change 2022, 170 (1), 1–26. https://ideas.repec.org/a/spr/climat/v170y2022i1d10.1007_s10584-021-03298-4.html
34. Sawtooth Software. Addressing Multicollinearity: Definition, Types, Examples, and More. sawtoothsoftware.com. https://sawtoothsoftware.com/resources/blog/posts/addressing-multicollinearity
35. Vani, G. Multicollinearity: Review of Problem and Its Diagnostics. rstudio-pubs-static.s3.amazonaws.com. https://rstudio-pubs-static.s3.amazonaws.com/919228_27a73663f18845fe8984fe47fbc79e60.html
36. Queensland Government. A Best Practice Guide for Local Governments Flood Classifications in Queensland. https://www.qra.qld.gov.au/sites/default/files/2020-06/flood_classifications_in_queensland_-_a_best_practice_guide_for_local_governments_-_may_2020.pdf.

Authors

Adrian Huihao Zhuang studies at The King's School Parramatta. His research on flood prediction was recognised at ISEF with a Third Place Award. Adrian interns at Hydgene Renewables, engineering diazotrophs to improve Australian soil. He ranks in the top 25 Earth & Environmental Science students in Australia. Contact at zhuanghuihao2006@gmail.com

Alex Huijie Zhuang studies at The King's School Parramatta. He was recognised at ISEF with a Third Place Award for his research in flood prediction. Alex interns at the Biomega Australia Company, engineering Omega-3 Fatty Acid-producing microbes. Alex is currently working in Earth and Environmental Science and Computer Science. Contact at zhuanghuijie2008@gmail.com.

Dr. Eric Sakk is an Associate Professor of Computer Science at Morgan State University. He has a Ph.D. in Electrical and Computer Engineering from Cornell University with a minor in applied mathematics. His background is in the field of dynamical systems, signal processing, computer systems communications systems, and bioinformatics. Prior to and during his post-doc, he was a lecturer in the Department of Electrical and Computer Engineering at Cornell University where he taught undergraduate and graduate courses in communications systems, signal processing and dynamical systems theory. His current research interests are in the fields of quantum computation, quantum communications, artificial intelligence and deep learning.