Human and Wildlife Coexistence in Canadian National Parks:

Discovering trends in reported incidents and identifying target areas for promoting health and safety of humans and wildlife and mitigating negative incidents in 35 Canadian National Parks

Emma Billard, Student Number: 501114915
Toronto Metropolitan University
CIND820 XJH – Big Data Analytics Project – W2023
Professor Tamer Abdou, PhD
February 20, 2023

Ryerson University

Table of Contents

Abstract

Time spent in nature is wondrous. Whether you're drawn to witness epic mountainscapes, giant old growth forests, or wildlife in their natural habitat, there is always wonder to be found in the wild. But we must not forget that being able to witness our beautiful natural world is a privilege and a gift. We must care for our natural planet so that it continues to thrive year after year.

In Canada, our National Parks are maintained by the Parks Canada Agency. According to Government of Canada (2022), the Agency's mandate includes acting as guardians of the national parks and protecting our natural places to ensure they remain healthy and whole. Under this mandate, Gummer & Nicholl (2022) indicate that between 2010 – 2021, Parks Canada compiled four datasets of incidents of human-wildlife coexistence in 35 National Parks for the evaluation of trends to inform Parks Canada policies and to ensure safe visitor experiences while conserving wildlife and integrity of our ecosystems.

For my data analytics project, I am using these four datasets, which contain 70,000+ records of incidents, responses, animals involved, and human activities related to human-wildlife coexistence. I combined these datasets using the unique Incident Number's associated with each record. I will not be referring to the thirteen other datasets included in the same Open Record and which contain compiled summaries of the incidents as I will be conducting my own pattern mining and summarizing.

The main problem I seek to address is to identify target areas for promoting health and safety of humans and wildlife and for mitigating negative incidents in our National Parks. I will be using the theme of predictive analysis, specifically pattern mining and causality. The goal is to determine which incidents are the most serious, and find patterns and correlations that will allow me to make predictions on when, where, and why these incidents occur. This information will

allow me to develop recommendations for park visitors and park employees to help ensure the health and safety of both humans and wildlife. I will tackle this by focusing on the following research questions:

1. What patterns can be found in location and time of year for each of the following variables: human activities, animals involved, cause, and type. How do these patterns differ year.
2. What incidents are the most concerning (i.e. where there is potential risk for humans or animals)?
3. What variables are most correlated with the occurrence of each incident type? Can we predict future similar incidents will occur near that location or that time of year?

Literature Review

Managing coexistence between humans and wildlife is something that communities have been dealing with globally for generations. In current times, the coexistent relationships come in many forms. Sometimes this might look like protecting livestock from wildlife, preserving wildlife's natural habitats to protect endangered and at-risk species, protecting wild spaces and animals from poachers, or protecting civilians who live on the border of natural spaces from being targeted by wildlife.

In my literature review, I read through several journals, articles, and research papers on the topic of human-wildlife coexistence and there was one quote in particular that stood out to me as describing the continuing need for more research and work in this field.

> "Human–wildlife conflict is increasing in both frequency and severity worldwide and will likely continue to escalate. Protected areas are increasingly becoming islands of habitat surrounded by seas of cultivation and development. Wildlife

> and humans increasingly compete for space, resources, and places to call home. Although ecosystem-based approaches (including the development of corridors between protected areas) offer improved long-term protection for many species from a biological perspective, they also involve extensive regional opportunities for interaction and conflict between local people and wildlife. Without properly addressing HWC in the effort to conserve wildlife and their habitat, conservation efforts will lose stability and progress, as well as the support of local communities." (Madden, 2010)

König et al. (2020) point out that "[h]uman–wildlife conflict (HWC) is a key topic in conservation and agricultural research. Decision makers need evidence-based information to design sustainable management plans and policy instruments." Bath & Ench (2003) also describe "[t]he chance to view wildlife draws millions of visitors each year to the national parks of North America. The combination of a large number of people and abundant wildlife leads to a variety of wildlife-human interactions."

It therefore comes as no surprise that Parks Canada has been gathering data about incidents relating to human-wildlife coexistence in 35 national parks for the purposes of evaluating trends and informing policies with the goal of ensuring safe visitor experiences and conserving wildlife and ecosystems. Ellis (2022) discusses this initiative in an article about Banff National Park saying that "Parks Canada's management efforts concentrate on improving public awareness and minimizing human-wildlife conflict, improving habitat quality, security, and connectivity where possible, and minimizing road and rail mortality"

We are so lucky to have so much natural space in Canada and I believe it is so important to preserve and protect the wildlife (plants, ecosystems, and wildlife) that we have. In my research, I've discovered there is much more work that is needed in this field as there are still several gaps in areas of the world that have been researched, as well as the data that is being collected in the areas that are being researched. It is so important to be getting a clear picture of the current situation and the data that is being collected by Parks Canada is a good start but we could go even further in the data we are collecting and areas around Canada that we are researching.

It is important to note that the "incidents" in my data are relating to coexistence, whereas many research articles I've found relating to this topic refer to HWC as Human and Wildlife **Conflict**. For clarity, human-wildlife coexistence is defined by König et al. (2020) as "a dynamic but sustainable state in which humans and wildlife coadapt to living in shared landscapes, where human interactions with wildlife are governed by effective institutions that ensure long-term wildlife population persistence, social legitimacy, and tolerable levels of risk". And human-wildlife conflict, is described by Madden (2010) as "… when the needs and behavior of wildlife impact negatively on the goals of humans or when the goals of humans negatively impact the needs of wildlife."

Many of the coexistence incident types in my dataset could also be considered a conflict incident, for instance "Human Wildlife Interaction"; however there are also several incidents included that would not consistent a conflict, for instance "Wildlife sighting". Seeing as there is so much cross-over between research being done for coexistence and conflicts, I have included sources which reference both types in my review.

Baral et al. (2021) conducted a related study analyzing the patterns and drivers of human-wildlife conflict "with the intent to better management [of] human wildlife conflict (HWC) and wildlife conservation in mid-hills outside protected areas of Gandaki province, Nepal". This study "investigated temporal, seasonal and spatial distribution of human casualties caused by wildlife attacks." Part of their methodology used multivariate logistic regression between the 11 independent variables and the dependent variable of human death or injury due to wildlife attack. They also used the chi-square test of independence to look at the association between variables, a generalized linear model to predict effect of variables on wildlife attacks, and "a priori candidate model and ranked them based on Akaike Information Criterion (AIC) values. Those models with lowest AIC values were considered the appropriate for explaining the wildlife attacks" (Baral et al., 2021). The results of this research were listed as:

> … we conclude that patterns of wildlife attacks on humans are influenced by spatial and temporal factors. Leopard was the major wildlife cause HWC followed by black bear. The conflict ranged from crop raiding, livestock depredation, traffic collision, property damage, transmission of diseases and human attacks. Among them, human attack was the most critical expression of HWC and needs addressing sensitively to liaison the local support for wildlife conservation. (Baral et al., 2021)

Another related study was conducted by Naha et al. (2020). This study looked to identify factors that caused leopards to attach livestock in western Indian Himalayan Region (IHR). Naha et al. (2020) describes their methodologies as examining seasonal and temporal variation in habitat types using the chi-square test. They also used various analytical approaches including the Moran's test for spatial autocorrelation, generalized linear models with binomial and poisson structures, priori candidate models ranked based on AIC, and used the predictor variables of the dominant models to calculate conditional inference (CTREEs). They describe "The GLM models computes probabilities of an event based on a logistic regression framework while the CTREE uses a machine learning classification approach and assigns values to predefined categories" (Naha et al., 2020). The "results suggest that landscape features are major predictors of livestock predation by leopards in IHR" (Naha et al., 2020).

Marchini et al. (2019) proposes using strategic planning with data science and modelling to inform a framework for policy makers and citizens regarding the "conflict-to-coexistence continuum". They propose the planning process begins with a situation assessment before proceeding to decision-making, implementation, monitoring and evaluation.

The data collected by Parks Canada regarding reported incidents of human and wildlife coexistence (HWC) in 35 national parks is the beginning "situation assessment" phase and my project aims to help complete the situation assessment and begin proposing actions for the decision-making phase.

In a personal communication (February 3, 2023), D. Gummer, Wildlife Management Specialist from Parks Canada, shared that so far, this data has been used to build an internal system that staff and management of the Parks can use to view and analyze incident data for their day-to-day operations. Parks Canada is interested in building a similar tool for the public, but that has not yet been created. "The data and ongoing analyses are also helping to inform new national policy/guidance that [they] are working on and many more that [they] propose for [the] future;" however there is not yet a public report that has been written to summarize the results of this data collection or how it has been applied in policies. Gummer shared that individual parks also use the data to develop their own local policies/reports, but there was no clear line from the data to the specific policies it has addressed.

While the data I am using is public, I have not been able to find any other usage of this exact data online.

Seeing as the analysis and results of this data that has been done to date has been not shared publicly, there is no way to know exactly what analysis has been done with this data. My research in important to discover and highlight trends in coexistence incidents over time and geographic locations and aims to find correlations between attributes that can predict incidents and identify target areas where improved protection is needed.

As someone who has worked for an Environmental charity for the past 7 years, this topic is very close to home for me. There is a constant push-and-pull between the desire to be in nature exploring and witnessing wildlife and the need protect and preserve these spaces. In fact, most people will say it’s the experiences they’ve had out in nature that have truly helped them understand the great importance of protecting the natural world. This is why it’s so important to ensure that we are able to balance these two needs – being able to enjoy the natural world, while also taking every opportunity given to us to protect it. We must continue strive to improve our coexistence with wildlife for all of our benefits.

Data Description

The combined dataset contains 73658 rows and 170 columns and has a combination of numeric (float64) and categorical (object) data. I’ve included some figures here that depict summary statistics and data types for the dataset I am using. I also have a complete Exploratory Data Analysis Report available (along with the Complete_HWC_Data.csv data) in my GitHub repository which can be accessed here: https://github.com/ebillard06/human-wildlife-coexistence-data-analysis.

Figure 1: Data Types

The data types for attributes in columns 0-19 are listed in Figure 1. The attributes in columns 20-170 are one-hot encoded columns generated from the categorical data from the “Activity Type” and “Response Type” attributes from the datasets before they were merged. All data types for columns 20-170 are float64.

```
UniqueID                              object
Incident Number                       object
Incident Date                         object
Field Unit                            object
Protected Heritage Area               object
Incident Type                         object
Latitude Public                      float64
Longitude Public                     float64
Within Park                           object
Total Staff Involved                 float64
Total Staff Hours                    float64
Species Common Name                   object
Sum of Number of Animals             float64
Animal Health Status                  object
Cause of Animal Health Status         object
Animal Behaviour                      object
Reason for Animal Behaviour           object
Animal Attractant                     object
Deterrents Used                       object
Animal Response to Deterrents         object
dtype: object
```

Figure 2: Numerical Data Descriptive statistics

Contains descriptive summary statistics for the attributes in columns 0-19 of the numerical data "float64" type.

```
Complete_HWC_Data[Complete_HWC_Data.columns[0:20]].describe()
```

	Latitude Public	Longitude Public	Total Staff Involved	Total Staff Hours	Sum of Number of Animals
count	73624.000000	73624.000000	73658.000000	73658.000000	73655.000000
mean	51.484498	-114.770930	1.477789	2.331069	2.728776
std	1.900213	9.784865	1.055412	14.361819	14.389458
min	41.902015	-140.297738	0.000000	0.000000	0.000000
25%	51.168223	-118.063343	1.000000	0.500000	1.000000
50%	51.286676	-116.165634	1.000000	1.000000	1.000000
75%	52.872448	-115.551471	2.000000	2.000000	1.000000
max	73.998028	-52.637169	32.000000	2400.000000	2000.000000

Figures 3 & 4: Categorical Data Descriptive Statistics

Combined, these figures contain descriptive statistics for the attributes in columns 0-19 of the categorical data “object” type. These figures are split in two to help with readability.

```
Complete_HWC_Data[Complete_HWC_Data.columns[11:20]].describe(include='object')
```

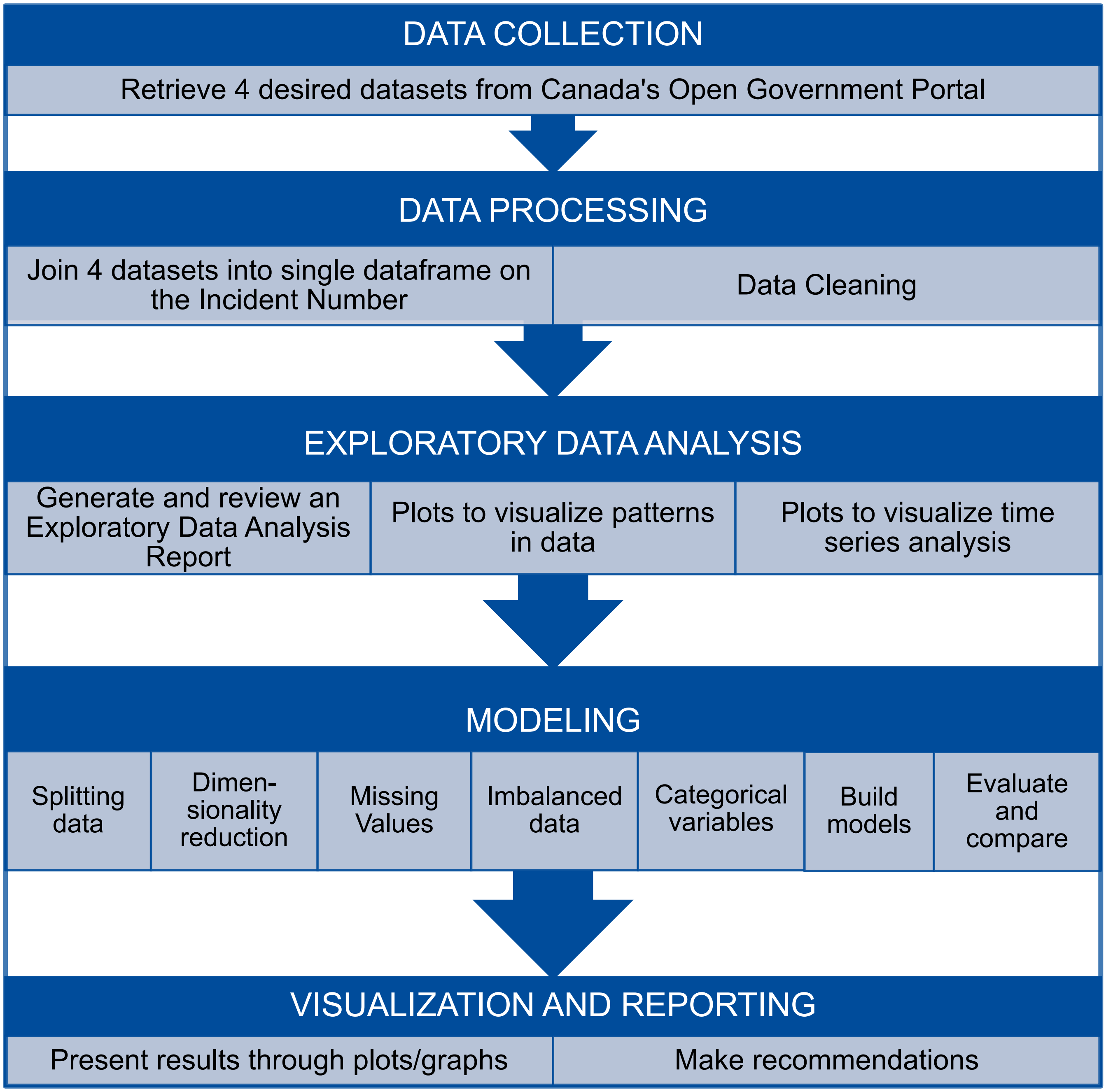
	Species Common Name	Animal Health Status	Cause of Animal Health Status	Animal Behaviour	Reason for Animal Behaviour	Animal Attractant	Deterrents Used	Animal Response to Deterrents
count	73655	41485	13197	45675	26138	24820	19544	10502
unique	322	9	17	23	17	21	26	10
top	Black Bear	Healthy	Collision	Presence - Wildlife Exclusion Zones	Habituation	Vegetation (natural)	Noise - Voice	Retreat - Run
freq	20898	25719	5752	17003	15559	9661	2474	4686

```
Complete_HWC_Data[Complete_HWC_Data.columns[0:11]].describe(include='object')
```

	UniqueID	Incident Number	Incident Date	Field Unit	Protected Heritage Area	Incident Type	Within Park
count	73658	73658	73658	73658	73658	73658	73618
unique	73658	64258	4299	19	35	9	2
top	BAN2010-0003.3	BAN2013-1151	2021-05-26	Jasper Field Unit	Banff National Park of Canada	Human Wildlife Interaction	Yes
freq	1	11	96	25982	27030	48673	72755

Project Approach (Methodology)

Figure 5: Graph of overall methodology.



Data Collection

Retrieve the 4 desired datasets from Canada's Open Government Portal: https://open.canada.ca/data/en/dataset/cc5ea139-c628-46dc-ac55-a5b3351b7fdf.

Data Processing

This phase has been started by dealing with duplicate "Incident Number" observations in each of the 4 datasets. In the "Animals" dataset, I created a "UniqueID" column based on the "Incident Number" value and a running count of occurrences of that "Incident Number" so each observation had a unique identifier. In the "Incident" dataset, each duplicate occurrence of the "Incident Number" held and NA value as the "Incident Type" (and held no other new information) so I simple dropped those rows. In the "Activities" and "Responses" datasets, the "Activity Type" and "Response Type" attributes respectively, were encoded using one-hot encoding so that each distinct category has a column with binary values of "0" for no and "1" for yes. I decided on this approach after discussing the options with Professor Abdou and I decided encoding these two variables was the best way to maintain the information in the dataset and be able to merge the 4 datasets together.

I then merged the four datasets together using the "UniqueID" value generated in the "Animals" dataset to ensure there is unique identifier for each observation in the newly combined dataset, named Complete_HWC_Data.

Next, I will do some data cleaning to deal with typos or mistaken entries. To find these errors I will compare unique values for "Activity Type", "Response Type", and other categorical attribute values with the values that are listed in Data Dictionary which was provided with the Open Data record to ensure valid entries.

Exploratory Data Analysis

Generate an Exploratory Data Analysis Report using the Panda's ProfileReport. Use report to observe variance within variables, correlation between variables, missing values, distinct values per variable, etc.

Use histograms and various other plots to subset data and explore patterns in incident types by human activity, animals involved, location, and time of year. Use statsmodels Python module and Time Series Analysis (tsa) to plot and visualize trends over time.

Modeling

Before deciding on which models to use, there are some preprocessing steps that needed to be looked into and addressed. Because these steps will be dealt with in the creation of the model, I’ve included this information as part of the Modelling phase of the analysis instead of part of the Preprocessing phase.

Splitting Data

Splitting data into training and test data for the purpose of predictive modelling. I will use TimeSeriesSplit function of the scikit-learn package to split and cross-validate my training and test data sets. As described by Howell (2023), doing a TimeSeriesSplit “means our test [data] is always in the future compared to the data our model is fitted on.”

Dimensionality Reduction

When researching dimensionality reduction techniques looking for the best fit for my dataset containing mixed data (both numeric and categorical data). Mahmood (2021) indicates that “[w]hen dealing with mixed data, FAMD is a recommended way to handle the unnecessary factor and reduce the dimensions of the data". Mahmood (2021) goes on to describe that Factor Analysis of Mixed Data (FAMD) analyzes the dataset using a combination of Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA) techniques to determine which attributes are the most critical components for modeling the data.

Dealing with Missing Values

The dataset has a few variables with a few (between 3-40) missing values. These variables are "Incident Type", "Species Common Name", "Sum of Number of Animals", "Activity Type" variables, "Latitude Public", "Longitude Public", and "Within Park". Because there are only a few missing values, I will impute this data using the KNN Imputation method.

The "Response Type" variables are missing values for around 1500 observations. This is still only about 2% of the data so I will also impute this data using the KNN Imputation method.

Lastly, there are some variables with several missing values (over 30,000). These include "Animal Health Status", "Cause of Animal Health Status", "Animal Behaviour", "Reason for Animal Behaviour", "Animal Attractant", "Deterrents Used", "Animal response to Deterrents". These attributes have too many missing values to consider any kind of imputation. I don't want to completely discount this data so I will still use it for pattern mining by sub-setting the data to all rows that have complete values for these variables and generate some graphs and correlation measures (using EDA report) to view it separately from the rest of the data. When it comes time to do modelling, these attributes will be dropped from the training data.

Dealing with Imbalanced Data

My dependent variable, Incident Type, is very imbalanced. Boyle (2019) succinctly addresses the problem with imbalanced classes as “Most machine learning algorithms work best when the number of samples in each class are about equal. This is because most algorithms are designed to maximize accuracy and reduce error.”

In my research into the pros and cons of the options for dealing with imbalanced classes, both under sampling and over sampling options have some downsides. Under sampling means you are not using some of the data from the majority class in order to balance the classes. In this dataset, the majority class contains about 66% percent of the observations, with the next highest class consisting of about 18% of the observations, so there could be a lot of valuable predictive information in the majority class that would be eliminated with under sampling. Over sampling adds data points (either randomly or synthetically generated) to the minority classes to balance the classes. Stewart (2020) points out one of the concerns with using over sampling is the risk of adding false data points could lead to overfitting. In this dataset, we would need to add very large amounts of datapoints to the minority classes in order for them to balance with the majority. That many “false” data points would cause concern that the data is no longer providing an accurate depiction of the actual data.

I also researched a third method of dealing with imbalanced the cost-sensitive learning method. Brownlee (2020) describes that “while most machine learning algorithms assume that all classes have an equal number of examples… [cost-sensitive learning] modifie[s] [algorithms] to change the way learning is performed to bias towards those classes that have fewer examples in the training dataset.” Stewart (2020), Brownlee (2020), and Selvan (2020) all claim that using cost-sensitive learning to deal with imbalanced datasets - where class weights are taken into account in the machine learning algorithm – perform more effectively than under or over sampling. I’ve decided to use this method to address the imbalance in my Incident Type variable.

Dealing with Categorical Variables

Most decision tree classifiers (including the commonly used scikit-learn decision tree classifier and random forest classifier) cannot handle categorical data without having it encoded to numeric form. When the categorical data is ordinal, this process is quite straight forward; however, with nominal data (where there is no inherit order), the process of converting categories to numeric involves creating “a new dummy variable for each level of the original variable” (Dingwall and Potts, n.d.). There are two main methods for encoding nominal categorical data: one-hot encoding and get_dummies. Kumar (2021) explains that the get_dummies method can be problematic if your test data set has a different set of columns depending on whether or not some categories exist in the test set that do or do not exist in the training set. The one-hot encoding method does not have this issue. Dingwall and Potts (n.d.) describe the following two potential problems that one-hot encoding can have in tree-based models:

1. The resulting sparsity virtually ensures that continuous variables are assigned higher feature importance.
2. A single level of a categorical variable must meet a very high bar in order to be selected for splitting early in the tree building. This can degrade predictive performance.

Even with those stated potential problems with using one-hot encoding, it is still a very popular and commonly used method. The alternative to encoding categorical data is of course leaving your categorical data as-is and using a tool like H2O for Python which can build a decision tree or random forest model using categorical data.

Based on my findings, I've decided to use one-hot-encoding on two of the three models I am generating using scikit-learn, and will maintain the categorical data for the last model using H2O. More specifics about these three models is included in the section below.

Models

I will be building three different models to compare the results. All three models will be using tree classifiers, but will have differences in the type of model or treatment of categorical variables. They are summarized below:

1. Decision tree classifier: using scikit-learn in Python and Gini index as the selection measure. For this model, one hot encoding will be applied to all categorical variables and I will use the sklearn.utils.class_weight function to balance the class weights.
2. Random Forest Classifier using scikit-learn: For this model, one hot encoding will be applied to all categorical variables and I will use sklearn.utils.class_weight function to balance the class weights.
3. Random Forest Classifier using H20: For this model, the categorical variables will remain as they are (with string values) and I will use the balance_classes option in H20 to balance the class weights.

Evaluation

Evaluate and compare the results of each model. The common evaluation metrics are accuracy, precision, recall, and F1 score. Accuracy is often not the best metric to use when evaluating imbalanced datasets. Precision is a good metric when your main aim is to minimize false positives and recall is a good metric when you want to maximize the true positive rate. In this situation, I see it as more important to maximize the true positive rate (successfully predict incident types) than it is to minimize false positive rates so while I will be considering all the performance metrics in my evaluation, I will use recall as my main performance metric.

Visualization And Reporting

I will present the results of the project using various plots and graphs. Based on the research I've conducted, I will make recommendations for the mitigation of incident by type with intention of promoting health and safety of humans and animals in these 35 National Parks.

References

Baral, K., Sharma, H. P., Rimal, B., Thapa-Magar, K., Bhattarai, R., Kunwar, R. M., Aryal, A.,& Ji, W. (2021). Characterization and management of human-wildlife conflicts in mid-hills outside protected areas of Gandaki province, Nepal. Plos One. https://doi.org/10.1371/journal.pone.0260307

Bath, A. J. & Enck, J. W. (2003). Wildlife-Human Interactions in National Parks in Canada and the USA. Social Science Research Review, 4(1), 1-32. http://npshistory.com/publications/wildlife/ssrr-v4n1.pdf

Boyle, T. (2019, February 3). Dealing with Imbalanced Data. Towards Data Science. https://towardsdatascience.com/methods-for-dealing-with-imbalanced-data-5b761be45a18

Brownlee, Jason. (2020, August 7). Multi-Class Imbalanced Classification. Machine Learning Mastery. https://machinelearningmastery.com/multi-class-imbalanced-classification/

Dingwall, N. & Potts, C. (n.d.). Are categorical variables getting lost in your random forests?. notebook.community. https://notebook.community/roaminsight/roamresearch/BlogPosts/Categorical_variables_in_tree_models/categorical_variables_post

Ellis, C. (2022, October 14). Parks Canada working to reduce growing human-wildlife conflicts. Rocky Mountain Outlook. https://www.rmoutlook.com/banff/parks-canada-working-to-reduce-growing-human-wildlife-conflicts-5920752

Government of Canada. (2022, November 19). *The Parks Canada mandate and charter*. https://parks.canada.ca/agence-agency/mandat-mandate

Gummer, D., & Nicholl, S. (2022, September 15). *Human-wildlife coexistence incidents in selected national parks from 2010 to 2021*. Government of Canada. https://open.canada.ca/data/en/dataset/cc5ea139-c628-46dc-ac55-a5b3351b7fdf

Howell, Egor. (2023, January 10). How To Correctly Perform Cross-Validation For Time Series. Towards Data Science. https://towardsdatascience.com/how-to-correctly-perform-cross-validation-for-time-series-b083b869e42c

König, H., Kiffner, C., Kramer-Schadt, S., Fürst, C., Keuling, O., & Ford, A.T. (2020). Special Section: Challenges of and Solutions to Human-Wildlife Conflicts in Agricultural Landscapes: Human-wildlife coexistence in a changing world. Conservation Biology, 34(4), 786-794. https://www.zoology.ubc.ca/conservation/wp-content/uploads/2021/04/Konig-et-al-2020.pdf

Kumar. S. (2021, August 11). Stop Using Pandas get_dummies() for Feature Encoding. Towards Data Science. https://towardsdatascience.com/stop-using-pandas-get-dummies-for-feature-encoding-5d2cd07cb4fc

Madden, F. (2010). Creating Coexistence between Humans and Wildlife: Global Perspectives on Local Efforts to Address Human – Wildlife Conflict. Human Dimensions of Wildlife: An International Journal, 9(4) 247-257, https://doi-org.ezproxy.lib.torontomu.ca/10.1080/10871200490505675

Mahmood, S, Md. (2021, July 11). Factor Analysis of Mixed Data. Towards Data Science. https://towardsdatascience.com/factor-analysis-of-mixed-data-5ad5ce98663c

Marchini, S., Ferraz, K. M. P. M. B., Zimmermann, A., Guimarães-Luiz, T., Morato, R., Correa, P. L. P., & Macdonald, D. W., (2019). Planning for Coexistence in a Complex Human-Dominated World Title of chapter. In A. Editor (Ed.), Human-Wildlife Interactions:

Turning Conflict into Coexistence (pp. 414-438). Cambridge University Press. https://doi.org/10.1017/9781108235730.022

Naha, D., Dash, S. K., Chettri, A., Chaudhary, P., Sonker, G., Heurich, M., Rawat, G. S. & Sathyakumar, S. (2020). Landscape predictors of human-leopard conflicts within multi-use areas of the Himalayan region. Scientific Reports: nature research, 10(11129). | https://doi.org/10.1038/s41598-020-67980-w

Selvan, T. S. (2020, August 31). How to handle Multiclass Imbalanced Data? – Say No To SMOTE. Towards Data Science. https://towardsdatascience.com/how-to-handle-multiclass-imbalanced-data-say-no-to-smote-e9a7f393c310

Stewart, Matthew. (2020, July 20). Guide to Classification on Imbalanced Datasets Towards Data Science. https://towardsdatascience.com/guide-to-classification-on-imbalanced-datasets-d6653aa5fa23