EcoAnalytics: Project Proposal

Vespa Velutina Wasp Nest Prediction

January 18th, 2021

+34 649 732 524
Calle Padre Jesus Ordoñez 1
28002 Madrid España

Executive Summary

The following is a proposal produced by the EcoAnalytics team towards the ecological problem of invasive species in the Basque Country. As a small but focused consulting firm, the team can leverage the power of a variety of data and translate insights into actionable steps forward. This proposal provides a contextual overview of the main problem at hand, including the team's understanding and first steps toward the solution. The paper delves into the factors affecting the Vespa Velutina wasps' nesting locations briefly and continues to describe the methods which will be used to most effectively predict the nest locations in the following year.

Problem Context

The Vespa Velutina wasp was accidentally introduced in Europe in 2004 and is negatively impacting Spain ecologically and economically. Apiaries are a very attractive food source for these wasps since there is a 70% concentration of honeybees and other similar species. Honeybees have a significant role in the ecosystem as they pollinate wildflowers and help cultivate crops benefiting humans economically The invasion of the Vespa Velutina wasp has caused the reduction of approximately 50% of beehives in many European areas, so to control this invasion two methods are being used: nest destruction and bait trapping.

Paths Forward

In order to eliminate the Vespa Velutina wasp, scientists need to find the location of the nest. There are some factors that determine the nesting sites, namely temperature, humidity, light intensity, shelter from the rain, and protection from the wind. These factors determine the preservation of the nest which is essential for the survival of the nest. Scientists have identified some patterns regarding nesting behavior. For example, nets are often located in trees near rivers, as water is a fundamental element for the construction of the nest. In addition, the presence of beehives is another important factor regarding the nesting sites. In practice, most nests were found near one or more apiaries.

Given the nesting behavior of Vespa Velutina wasp, the team decided to use four datasets to build predictive models predicting the number of nests by the municipality in Bizkaia. The four datasets are Mapa Forestal CAE, Euskalmet Info, Beekeeping, and Nest. These four datasets contain all the necessary information determining the nesting sites.

The teams have decided to merge four datasets together by the municipality. Descriptive analysis can be done and insight information can be extracted from the combined master table. The team decided to define the first three datasets as X and define the last dataset as Y in order to build a prediction model predicting the number of nests by the municipality in Bizkaia. X factors are attributes of the municipality which includes types of terrain, tree, area, precipitation, temperature, frosty days, medium humidity, average daily irradiation, average wind speed captured, number of farms, number of beehives. Y is the number of nests found

in the municipality. The team intends to use supervised machine learning algorithms. Given the numeric nature of the demanded result, the team intends to use regression algorithms.

Project Outline

The team decided to organize this project following the CRISP-DM model. CRISP-DM stands for **CR**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining. It is a process with six phases that naturally describes the data science life cycle. The six phases are business understanding, data understanding, data preparation, modeling, evaluation, deployment. The team adjusted the six phases slightly according to the project's needs when scheduling for timelines. The schedules are shown in the table below.

Phase	Date	Deliverables
Business Understanding	January 15th - January 21st	Project Proposal
Data Understanding	January 22th - March 12th	Master Table, Descriptive Analysis
Data Preparation, Modeling, Evaluation	March 13th - March 25th	Clustering and Predictive Model
Final Presentation	March 26th	Final Presentation

The deliverable for the business understanding phase contains a project proposal where the team demonstrates understanding of the problem, proposes possible paths forwards, outlines project schedule and details final solutions. The deliverable for the data understanding phase includes the combined master table and descriptive analysis done on the master data. Descriptive analysis will be delivered in PowerPoint presentation and mainly focusing on patterns identified regarding nesting sites and visualization of the nesting sites. The data preparation, modeling and evaluation phases will deliver a predictive model which will be able to predict the number of nesting sites per municipality in Bizkaia province. The final presentation will include the descriptive analysis, the predictive model, and the dashboard.

Deliverables

The final product handed over to the Basque Government will consist of a cloud-hosted model, deployed. From prior research there were some interesting metrics which we will try to reproduce and improve further if possible. An example of this was to track the number of queens produced per nest - a vital variable to record as the queens are responsible for making their next destination thrive with their breed. Once trained on the relevant historical data, simply plug in the required data fields for the year of data and the model will churn out new predictions based on the newly entered data. Choosing the time-frame came down to looking at the situation - Vespa Velutina follows a similar life-cycle, allowing a forecasting timestep

to be used. The specific seasonal habits of the Vespa Velutina aids us in simplifying the problem by using discrete-time models to generalize their behavior.

With the predicted locations of the Vespa Velutina wasps of next year, the environmental workers can explore insights, but a dashboard will unlock the full potential of the data. Using PowerBI a dashboard will be developed with monitoring services of key metrics. The dashboard will include valuable insights mined from the datasets. Through the development process the product will evolve through several versions to ensure the final product is the best solution.