

Netwatch Sidekick

Masters of Data Science research project
in collaboration with Netwatch Ireland/UK

Authors

Derry Brennan
Researcher

Dr Greg Doyle
Supervisor

Affiliations



Institiúid Teicneolaíochta Cheatharlach



Related literature

Singh, A., N., M. and Lakshminathan, R. (2017) 'Impact of Different Data Types on Classifier Performance of Random Forest, Naive Bayes, and K-Nearest Neighbors Algorithms', International Journal of Advanced Computer Science and Applications, 8(12). doi:10.14569/IJACSA.2017.081201.

Raschka, S. (2020) 'Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning', arXiv:1811.12808 [cs, stat] [Preprint]. Available at: <http://arxiv.org/abs/1811.12808> (Accessed: 20 March 2022).

Wirth, R. and Hipp, J., 2000, April. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1, pp. 29-40).

Introduction

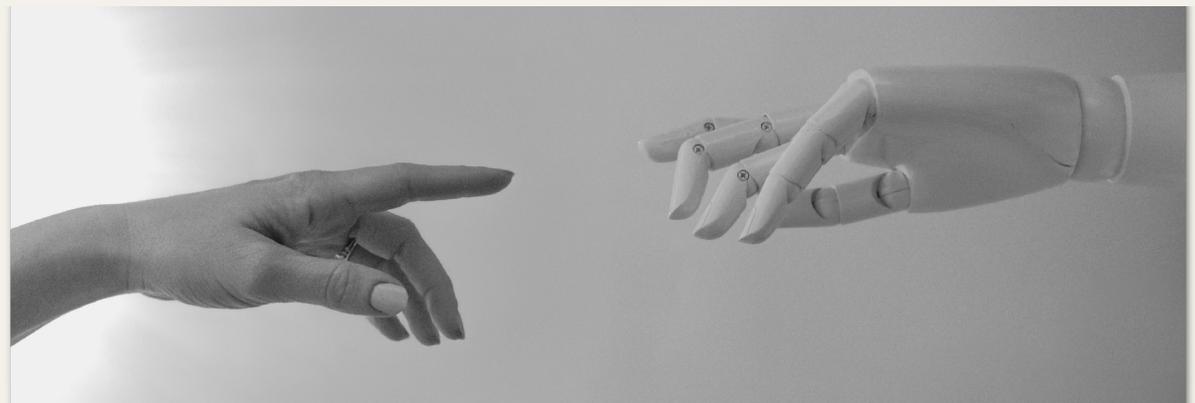
This project aims to create a recommendation engine ("sidekick") that will help human operators make better decisions on which action to take, based on past decisions made by humans in similar circumstances. A large number of circumstances should be taken into consideration. The machine learning model is based on the following inputs

- Alarm records (date, time, camera, GPS coordinates)
- Specialist response (go live, audio, call, isolation)
- Weather data

The model is built from years worth of data and is constantly updated with live data thereafter. The model aims to predict the most likely specialist response to each new alarm that arrives:

- Go live
- Audio
- Call
- Isolate

The model will ultimately assist inexperienced specialists in choosing between the options that they have by suggesting (based on historical data) a course of action. The specialist can ignore or follow the recommendation.



Objective

An alarm system is a piece of hardware (server) installed on a customer site that has cameras connected to it. The alarm system analyses the camera stream and when it detects motion over a certain threshold it sends an alarm record and image-set (3 still images) to a server in the cloud. The image set is stored in the cloud and the alarm record with the details (date, time, camera number, site reference) is stored in a database table. This table acts as a queue for specialists (operators) monitoring the alarm records.

The specialist uses a desktop app to pick the top item from the queue and examine the image set. Items on the queue that came from the same customer site are grouped into an alarm group so that the one specialist deals with all alarms from that one customer site at the same time. If the image set contains something of concern, the specialist will go live to investigate. They may then give audio (shout at an intruder) and place a call to the customer or police.

If the image-set shows something considered as a nuisance (e.g. a spider, a dog) and it has happened repeatedly, the specialist may create an isolation for a period of time which causes subsequent alarm records to bypass the queue and get stored directly into an archive table.



Methodology

Using alarm data provided by Netwatch Ireland/UK, in combination with Met Eireann's historic weather statistics supply the data to selected machine learning algorithms. These algorithms are random forest, K nearest neighbour and naive Bayes (Singh, N. and Lakshminathan, 2017).

The Sidekick project will be conducted using CRISP-DM. This data science model is widely used in processing and provides the framework that will allow for a well-structured approach to the problem (Wirth and Hipp, 2000).

The aforementioned clustering algorithms will be benchmarked on their predictive performance and use hyperparameter tuning to optimise their output (Raschka, 2020).

The training data includes GPS coordinates which have been anonymised by using an offset from actual positions to maintain the relative area while eliminating the exact position of the site.

Technologies Utilised



Next Steps

Train and test numerous deep learning algorithms and evaluate their accuracy.

Merging additional weather data from Met Eireann with Netwatch Alarm Data to provide atomic hour-level weather conditions for the year 2021.

Establish a relationship between data points and their influence on the alarm response.

Answer the research question:

Which machine learning algorithm out of random forest, k nearest neighbour and naive bayes provides the greatest accuracy in predicting the Netwatch specialist response and which attributes from the data are the best indicators of that response?