Data Science For Social Good
Summer Fellowship
The Alan Turing Institute
WARWICK
THE UNIVERSITY OF CHICAGO

Improving Outcomes for Rough Sleepers in the UK

homeless link

Lushi Chen
Fellow
University of Edinburgh

Zoë Kimpel
Fellow
Northwestern University

Austin Nguyễn
Fellow
Williams College

Harrison Wilde
Fellow
University of Warwick

Josh Sidgwick
Project Manager
The Alan Turing Institute

Adolfo De Unanue
Technical Mentor
University of Chicago

Gareth Thomas
Partner Team Lead

INTRODUCTION

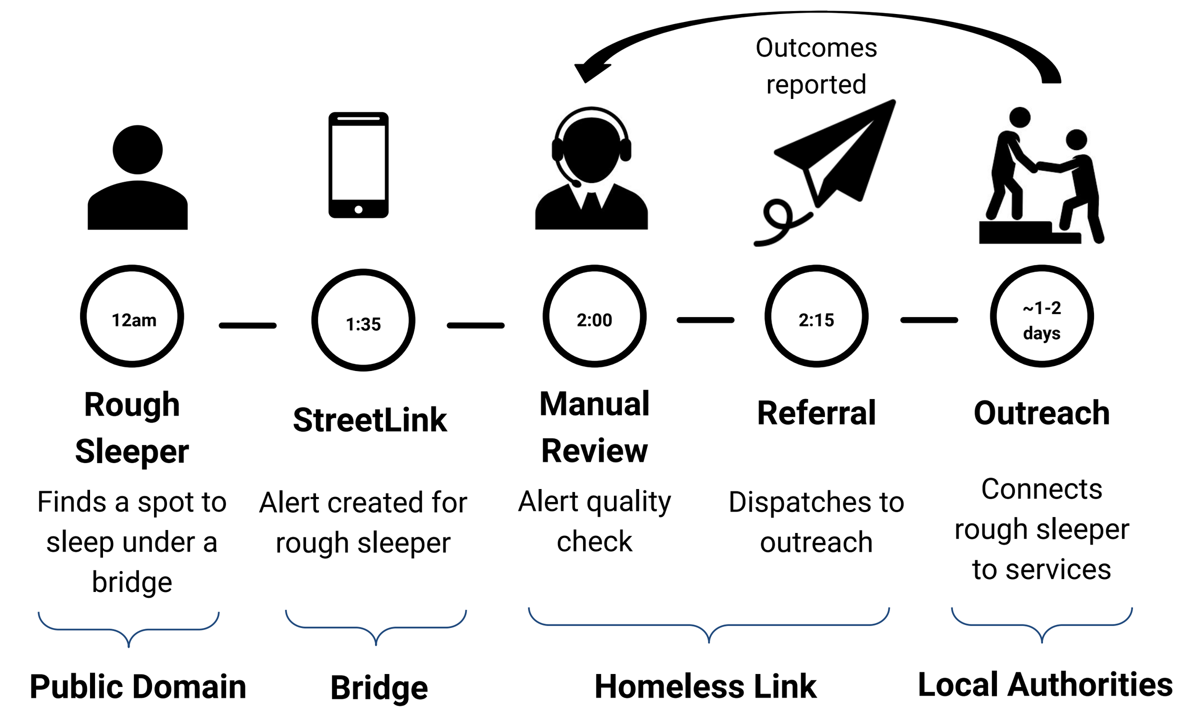
320k are homeless in England

4788 are rough sleepers

14% of alerts lead to finding a rough sleeper

Homelessness is a serious and growing issue in the UK. **Rough sleepers** are those who have no choice but to sleep on the street, leaving them amongst the most vulnerable of the homeless population. Homeless Link, a UK charity, works to help rough sleepers through StreetLink, a referral platform that connects them to essential services.

CURRENT PROCESS



280k ALERTS → **168k** REFERRALS → **39k** PERSONS FOUND

GOALS

We use our models to **generate a prioritised list** of incoming alerts based on their predictions to allow StreetLink to dispatch high quality referrals more quickly.

Faster referral dispatch will **connect more rough sleepers** to services in conjunction with rapid response teams.

Our exploratory analysis and models help to **quantify the rough sleeper issue** to encourage positive policy change in the future.

METHODOLOGY

data

We use internal StreetLink alert data from 2017 to 2019.

DEMOGRAPHICS
LOCATION
LOCAL AUTHORITY
FREE TEXT FIELDS
TEMPORAL
WEATHER

extract-load-transform

EXTRACT STREETLINK ALERT DATA → LOAD INTO DATABASE → TRANSFORM & CLEAN

feature engineering

LOCAL AUTHORITY CAPACITY: How often did a local authority find a person last week?

SPATIAL + TEMPORAL How many alerts have occurred in the past week at a given location?

TEXTUAL What was the rough sleeper wearing? Where is the rough sleeper located?

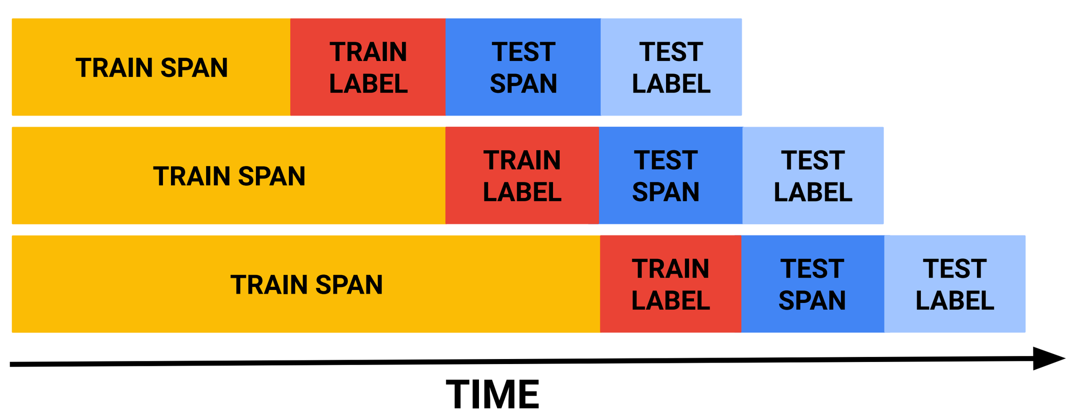
pipeline & experiments

We built a machine learning pipeline to run experiments with various classification algorithms such as **Random Forest, Gradient Boosting, KNN,** and **Decision Trees,** with different combinations of hyperparameters and feature sets.

Each model returns an output a score for each alert indicating alert quality.

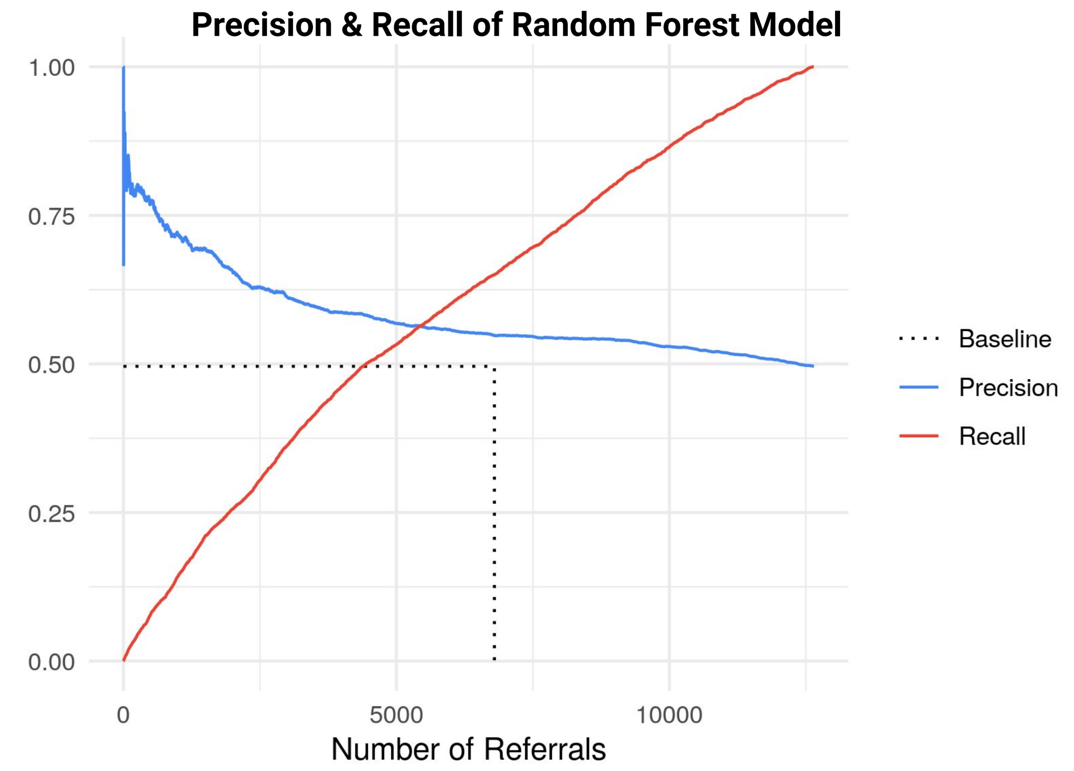
validation

Temporal cross-validation allows us to test and validate model performance over time.

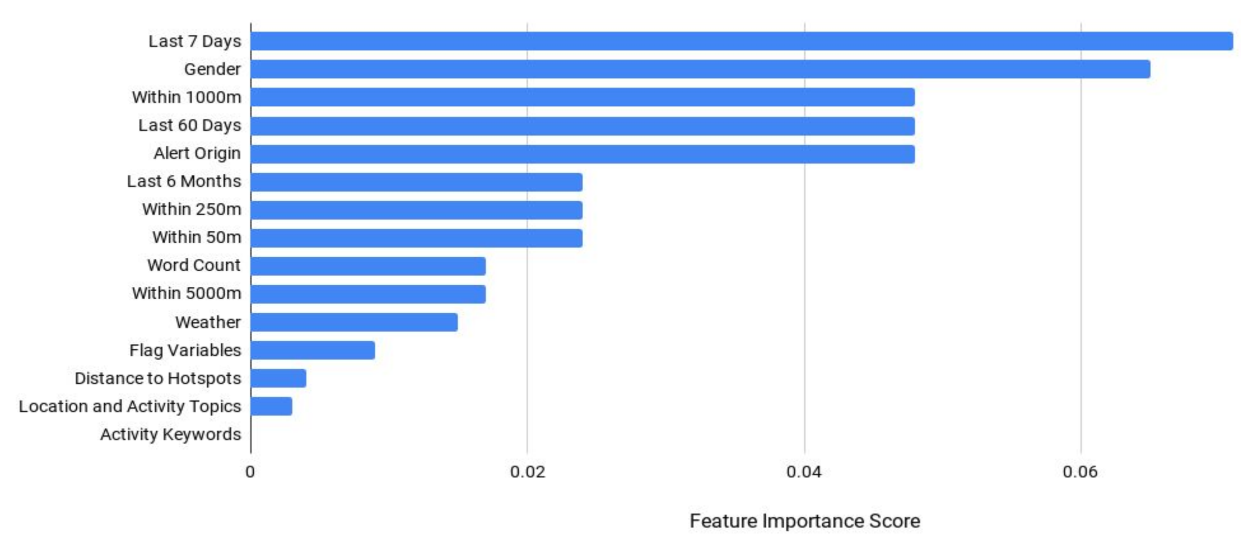


model selection

We selected a Random Forest model that maximizes precision at k and recall at *k*, where *k* is defined by StreetLink's resource constraints. We selected *k = 4000*.

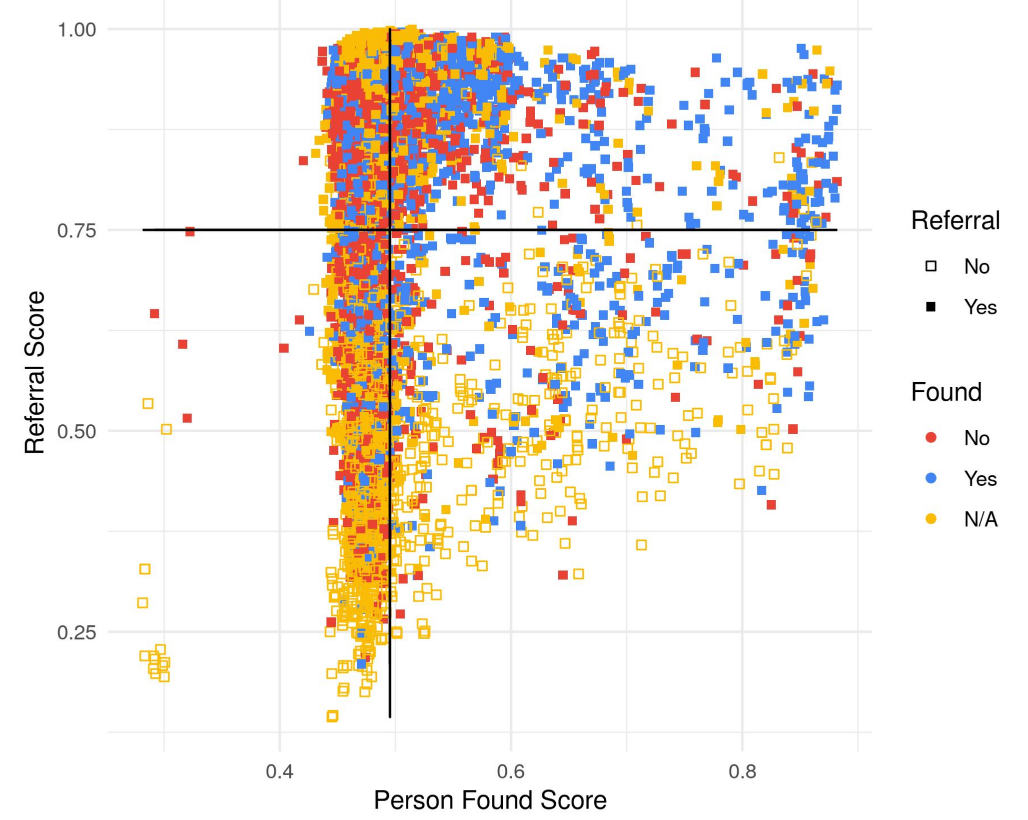


feature importance



This shows the highest feature importance score in each feature group from the random forest model. From 264 features, we grouped the features. **Recent temporal data** and **demographic** features were shown to be the most important in determining the likelihood a person would be found.

model interpretation



Our model prioritises high quality alerts and **challenges any existing biases** in StreetLink's current process.

Compared to low scoring alerts, the highest scoring alerts chosen by our model are:

23% farther away from hotspots on average
35% more cases of time seen
85% fewer alerts by phone

results

{ **18%** }
INCREASE IN ROUGH SLEEPER FOUND RATE WITH OUR MODEL

IMPACT

Better connect rough sleepers to services

Reduce time a person is forced to sleep rough on the streets

Advocate for policy changes to help rough sleepers

Thank you to our partners at Homeless Link and the StreetLink Staff for sharing their time, data and knowledge with us.

This work was done during the Data Science for Social Good Summer Fellowship at the University of Warwick and The Alan Turing Institute, hosted in partnership with the University of Chicago and the DSSG Foundation.