# INCORPORATING SPACE AND TIME INTO RANDOM FOREST MODELS FOR ANALYZING GEOSPATIAL PATTERNS OF DRUG-RELATED CRIME INCIDENTS IN A MAJOR U.S. METROPOLITAN AREA

<https://www.sciencedirect.com/science/article/abs/pii/S0198971521000065>

* objective: analyze changing patterns of drug-related crime incidents (heroin and synthetic drugs) in Chicago metropolitan area
* incorporating space and time
  + spatial and temporal lag variables added to detect dependencies and relationships
  + temporal granularity determined using clustering and time series analysis
  + moving average applied for smoothing and trend detection
* feature selection
  + key location identified for heroin and synthetic drug-related crimes
  + seven variables created based on alleys, vacant buildings, vacant lots, parking lots, gas stations, parks, and high schools
* model training and variables
  + random forest model trained with sociodemographic and built environment factors
  + variables include two lag variables, a trend variable, and key location features
* model optimization
  + guided regularized random forest used for feature selection
  + class imbalance handled using oversampling and underdamping techniques
  + grid search performed for hyperparameter tuning
* variable importance
  + corrected impurity importance used to measure contributions of variables
  + partial dependence plots constructed for spatial and temporal lag variables
* out-of-sample testing
  + model validated using out of bag error rate during training
  + independent test dataset used to assess forecasting performance for the last time periods of 2019
* results
  + OOB error rates for heroin and synthetic drugs indicate strong model performance (2.41% and 4.52%)
  + Key variables identified for hotspot classification include vacant lots, vacant buildings, sociodemographic factors, and specific key locations
* Space time random forest model effectively captures spatiotemporal patterns in drug related crimes and demonstrates strong predictive capabilities

# Predicting Crime Using Time and Location Data <https://dl.acm.org/doi/abs/10.1145/3348445.3348483>

* Objective: develop predictive models that can effectively analyze crime patterns based on both temporal and spatial factors, ultimately contributing to the understanding and anticipation of criminal activities.
* uses Chicago police departments CLEAR system of crime from 2001 to 2017
* employs random forest, decision tree, and ensemble methods like boosting and adaboost to predict crime based on the provided dataset.
* Location description: block, latitude, longitude
* Time: year, month, day, hour, minute, second
* Preprocessing steps involve handling string attributes, converting text features into numeric values, and splitting the ‘date’ attribute to extract relevant time-related features

# Crime Prediction Using Spatio-Temporal Data

<https://link.springer.com/chapter/10.1007/978-981-15-6648-6_22>

* Objective: the study aims to prevent and solve crimes by employing data driven research, acknowledging that a significant portion of crimes is committed by a small number of offenders
* San Francisco data over 12 year period
* Preprocessing
  + Training and testing dataset
    - To avoid overfitting and ensure realistic accuracy, dataset is divided into training and testing sets
    - Test set to 25%
  + Feature extraction and selection
    - Dividing day into distinct parts (early morning, late morning, afternoon, night)
    - Principal component analysis: used for linear dimensionality reduction, projecting data into a lower-dimensional space while maximizing variances
    - Sklearn’s feature selection module to identify the most relevant features for classification
* Methodology
  + Decision tree and k-nearest neighbor initially applied for crime prediction
  + Random forest and adaboost introduced to enhance prediction accuracy
    - Accuracy and log loss improve as the number of trees in the forest increases
* Performance measurement
  + Log-loss is used to evaluate classifier performance, considering the datasets highly imbalanced nature

# Mining 911 Calls in New York City: Temporal Patterns, Detection and Forecasting

<https://wrap.warwick.ac.uk/78529/>

* Objective: understand typical temporal behavioral patterns for emergency calls, considering various factors such as call type, location and day of the week
* Feature extraction
  + Primarily used historical means and medians for ICAD 911 and 311 complaint calls
  + Calculated over retrospective windows, considering day and time variations
* Data integration
  + Incorporated autoregressive features from past 911 and 311 data
* Each sample includes crime call count in a sector over an 8 hour period
* Feature reduction
  + Initial 2600 features reduced to just over 300
  + Pruned features with low variance and zero counts
  + Utilized feature importance from random forest regression
* Model implementation
  + Utilized a rolling forecast method for training and testing
  + Incorporating historical data incrementally, training on a sufficient initial dataset
* Random forest regression
  + Used mean squared error for individual trees
  + Employed 100 trees for the ensemble model
  + Poisson regression modeled the expected call-crime count of each sector using log-linear model
* Performance
  + Predicted call counts for entire days and 8 hour officer shifts
  + Evaluation metrics: RMSE, R2 and correlation coefficient
  + Comparison of rolling forecast model with extended forecast and Poisson regression models

# Predicting Violent Crime Reports from Geospatial and Temporal Attributes of US 911 Emergency Call Data

<https://arrow.tudublin.ie/scschcomdis/190/>