Crowdcast April 5, 2021 – Predict Wildfires in Australia

9am US Pacific, noon US Eastern, 6pm Central Europe; 8pm UAE, 9:30pm India

https://www.crowdcast.io/e/predicting-australian/

malaika@us.ibm.com Susan Malaika

Call for Code Spot Challenge for Wildfires Announcement on Nov 12, 2020

- Nearly 3 billion animals were affected by Australia's worst wildfire season that burned from July 2019 through March 2020 estimates Chris Dickman, a professor of ecology at the University of Sydney. The human cost to Aboriginal and Torres Island Australians, who lost their homes and their sacred sites, is devastating.
- Join data scientists to develop models focused on forecasting wildfires in Australia for the upcoming wildfire season and enter the chance to win 5K US Dollars. To get you started we're releasing historical data sets extracted from Weather Operations Center Geospatial Analytics component (PAIRS Geoscope) Our goal is to better understand the application of machine learning techniques in this domain.

- **Useful Links**
- The Contest landing page (this page) http://ibm.biz/cfcsc-wildfires
- The Contest GitHub https://github.com/Call-for-Code/Spot-Challenge-Wildfires
- The Contest leaderboard http://ibm.biz/cfcsc-wildfires-lead
- Slack Workspace http://callforcode.org/slack Channel #cfcsc-wildfires
- Helpful blog https://medium.com/ibm-data-ai/predicting-australian-wildfires-with-weather-forecast-data-8d1cc983c863

What Happened – 4 phases

Time-series problem based on daily data provided by Pairs Geoscope

Five datasets provided:

- Historical wildfire
- Historical weather
- Historical weather forecasts
- Historical vegetation index
- Land classes

Data Refresh	Submissions Take Place	Contest Stage	Max Allowed Number Of Submissions
Available on 2020-11-10 Base Data - starts between 2005 - 2015 until 2020-10-31	2020-11-10 ~~2021-01-09~~ 2021-02-28	Development - Try the platform - Predict Feb 2020 (the first 28 days)	daily 10, weekly 50, total 100
Available on 2021-01-09 Refresh data to include up until 2021-01-08 (with some exceptions - see notes in zip file)	2021-01-10 ~~2021-01-15~~ 2021-01-19	Predict Jan 2021 week 3 (Jan 16-22)	daily 5, weekly 35, total 35
Available on 2021-01-18 Refresh data to include up until 2021-01-14 (see zip files)	2021-01-16 ~~2021-01-22~~ 2021-01-26	Predict Jan 2021 week 4 (Jan 23-29)	daily 5, weekly 35, total 35
Available on 2021-01-23 Refresh data to include up until 2021-01-22 (see zip files)	2020-01-23 ~~2021-01-31~~ 2021-02-02	Predict Feb 2021 (Feb 1-28)	daily 3, weekly 3, total 3

Goal: Predict the size of the fire area in km squared by region in Australia for each day in February 2021 using data available up to January 29th

Call for Code Spot Challenge for Wildfires - Predict February 2021 Australia

The Contest leaderboard http://ibm.biz/cfcsc-wildfires-lead

Leaderboard after evaluation on Mar 2 with Feb data including Feb 28

Participant Team	Method Name	GitHub Url	rmse	mae	tot	Last submission at
Data Warriors	N/A	N/A	19.6021	7.031	9.5452	35 day(s) ago
Schildkröte	DNN, Linear Log, 16	N/A	21.0631	7.3531	10.0951	30 day(s) ago
dss	Feb2021_ak	N/A	21.0401	8.0493	10.6474	27 day(s) ago
MLVV	N/A	N/A	20.0402	9.0974	11.286	27 day(s) ago
TomasKaminskas	N/A	N/A	19.747	10.2913	12.1824	34 day(s) ago
OborITERA	LMU	N/A	20.6271	11.0571	12.9711	27 day(s) ago
Student	N/A	N/A	24.9991	16.9296	18.5435	27 day(s) ago
Katharina Knappmann	N/A	N/A	47.4495	19.3634	24.9806	27 day(s) ago
DNAwarriors	Sub2	N/A	48.0531	25.7235	30.1894	27 day(s) ago
MM	N/A	N/A	68.3319	29.5657	37.319	28 day(s) ago
NoSmoke	Watson - AutoAI	N/A	58.4924	33.5799	38.5624	27 day(s) ago
YHA	DNN	N/A	62.4899	49.2962	51.935	27 day(s) ago
Team-Bul	N/A	N/A	111.67	46.8967	59.8513	27 day(s) ago
titanio01001	N/A	N/A	169.1759	46.7604	71.2435	26 day(s) ago
ROTKK	N/A	N/A	145.94	74.6057	88.8726	28 day(s) ago

The Final External Leaderboard

Lowest Tot is good

Closest to reality

Crowdcast April 5, 2021 – Predict Wildfires in Australia

- Welcome and Call for Code Overview (Upkar)
- Overview & Summary Results (Susan)
- The datasets (Hendrik)
- A technical overview of the models from 3 teams
  - yau_yee_Italy Team (Marco)
  - Data Warriors (Albert)
  - NA Team (Brianna, Ned)
- A summary comparison of the modeling approaches (Wiktor)
- Students using the challenge as part of their university program
  - Dr Manar, Emad, Abdelhady
- What's next (Susan)
- Call for Code Close and Q&A (Upkar)

Crowdcast April 5, 2021 – Predict Wildfires in Australia

9am US Pacific, noon US Eastern, 6pm Central Europe; 8pm UAE, 9:30pm India

https://www.crowdcast.io/e/predicting-australian/

Call for Code Spot Challenge for Wildfires

The datasets

Comparing prediction models from 3 leading teams

Case study at university

Do join us https://www.crowdcast.io/e/predicting-australian/

More info http://ibm.biz/cfcsc-wildfires

Congratulations Data Warriors

Albert Um

Shruti Chaturvedi

Divyansh Choubisa

Call for Code Spot Challenge for Wildfires with PAIRS Geoscope

Hendrik F. Hamann, Johannes Schmude
IBM T.J. Watson Research Center

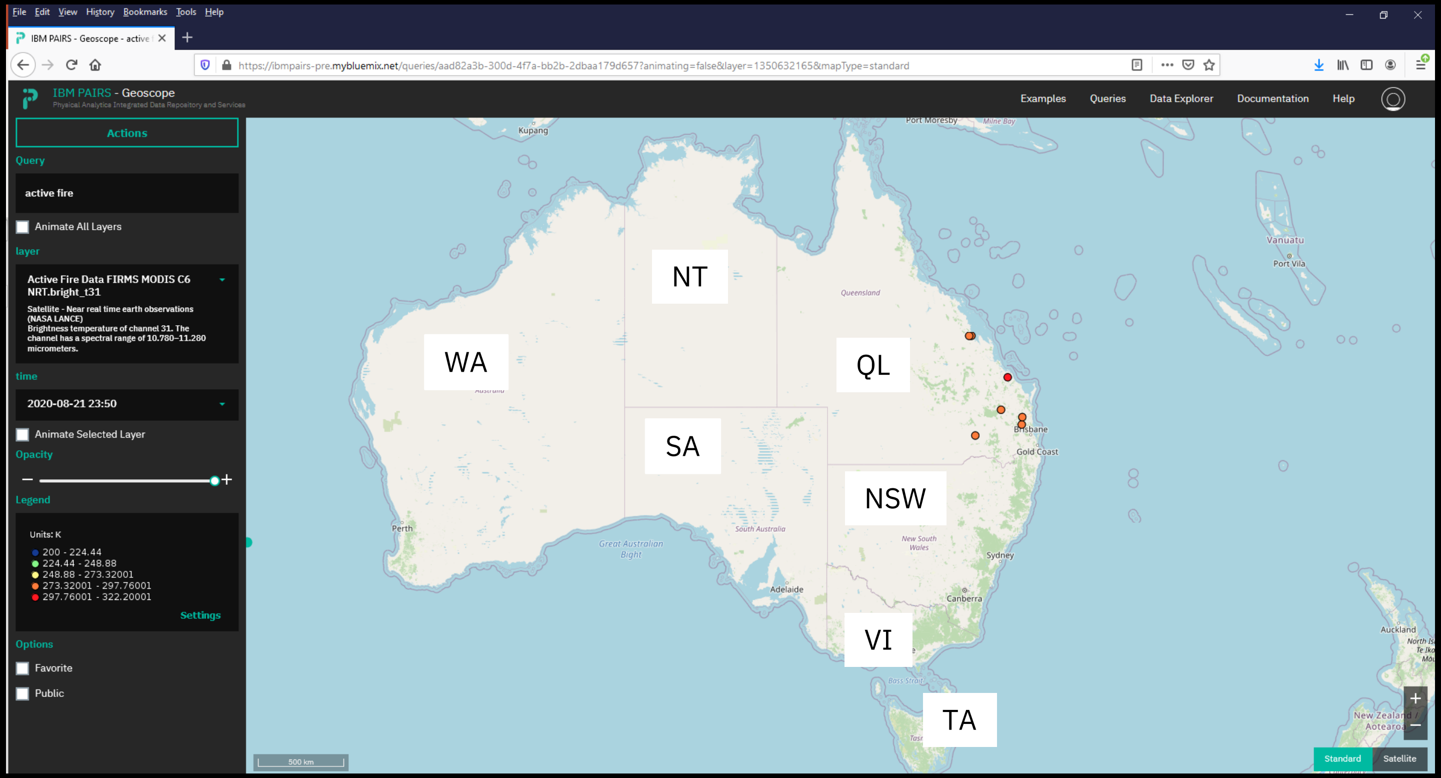
IBM

What will be forecasted/predicted in this contest?

Forecast/predict the daily total estimated fire area for 7 states in Australia for Feb. 2021?

NSW=New South Wales*
NT=Northern Territory
QL=Queensland
SA=Southern Australia
TA=Tasmania
VI=Victoria
WA=Western Australia

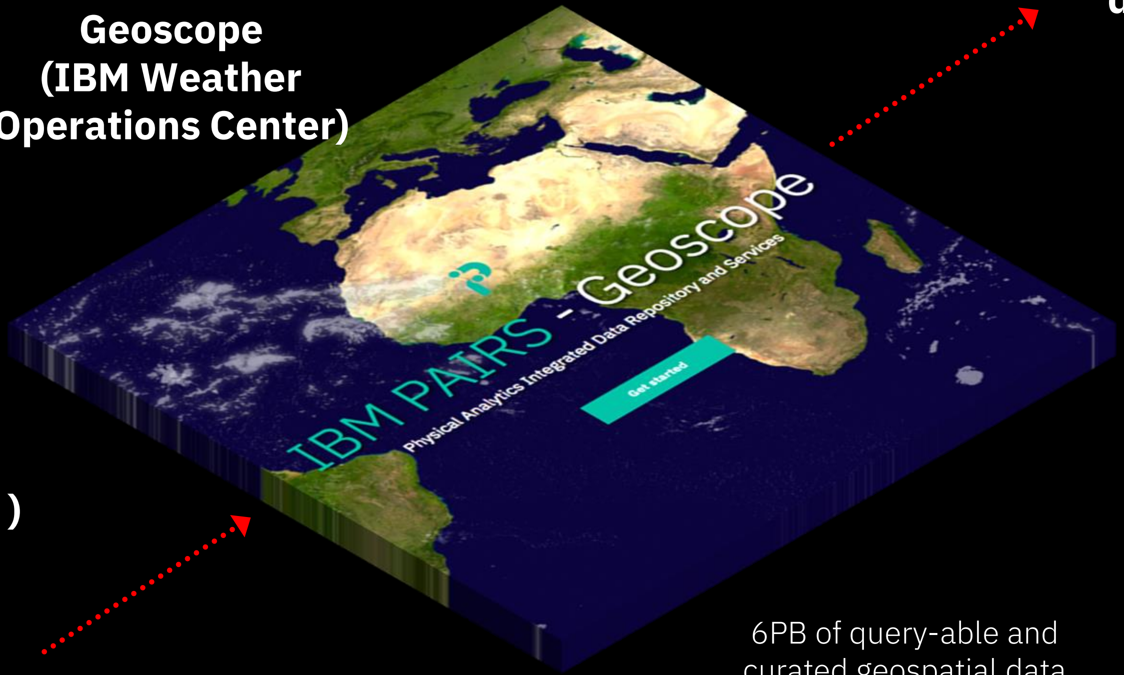
*excluded the capital region



© 2018 IBM Corporation

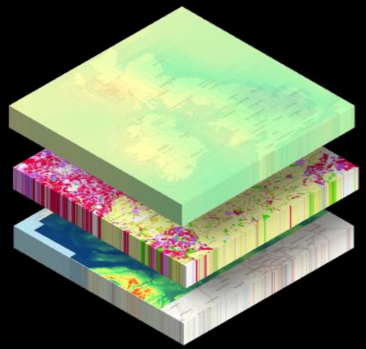
Call for Code Spot Challenge for Wildfires data sets have been "extracted" from PAIRS

IBM PAIRS Geoscope (IBM Weather Operations Center)



Extract for data science contest

Raw data (NASA, Copernicus, NOAA etc.)



6PB of query-able and curated geospatial data

© 2018 IBM Corporation

The extracted data

1. Historical Wildfires
2. Historical Weather
3. Historical Weather Forecast
4. Land Class
5. Normalized Vegetation Index

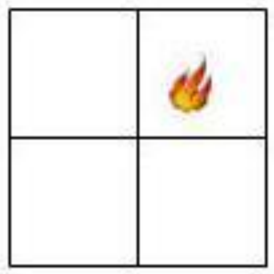
- Readme_Docs.docx V2
- HistorialWeatherForecasts.csv V2
- HistoricalWeather.csv V2
- VegetationIndex.csv
- LandClass.csv V2
- Historial_Wildfires.csv

© 2018 IBM Corporation

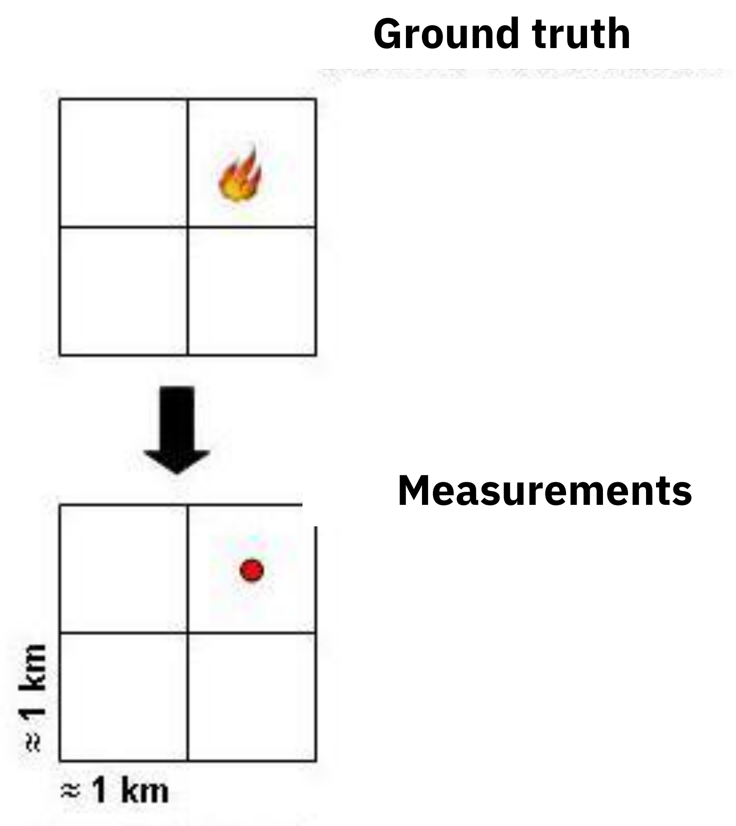
Historical fire data from a NASA satellite

Historical fire data – from hotspots to pixels to estimated areas

Ground truth

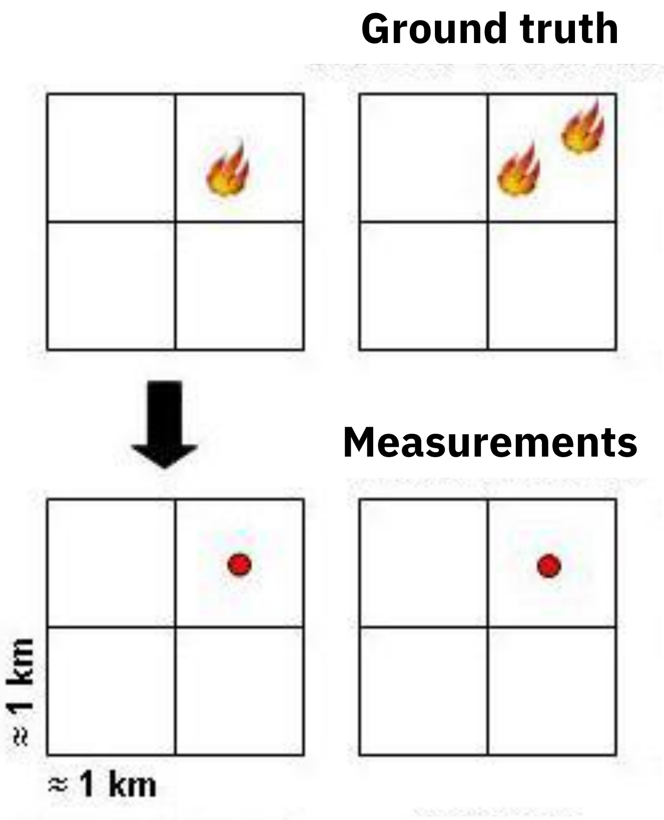


Historical fire data – from hotspots to pixels to estimated areas



$1\ km^2$
(1 fire)

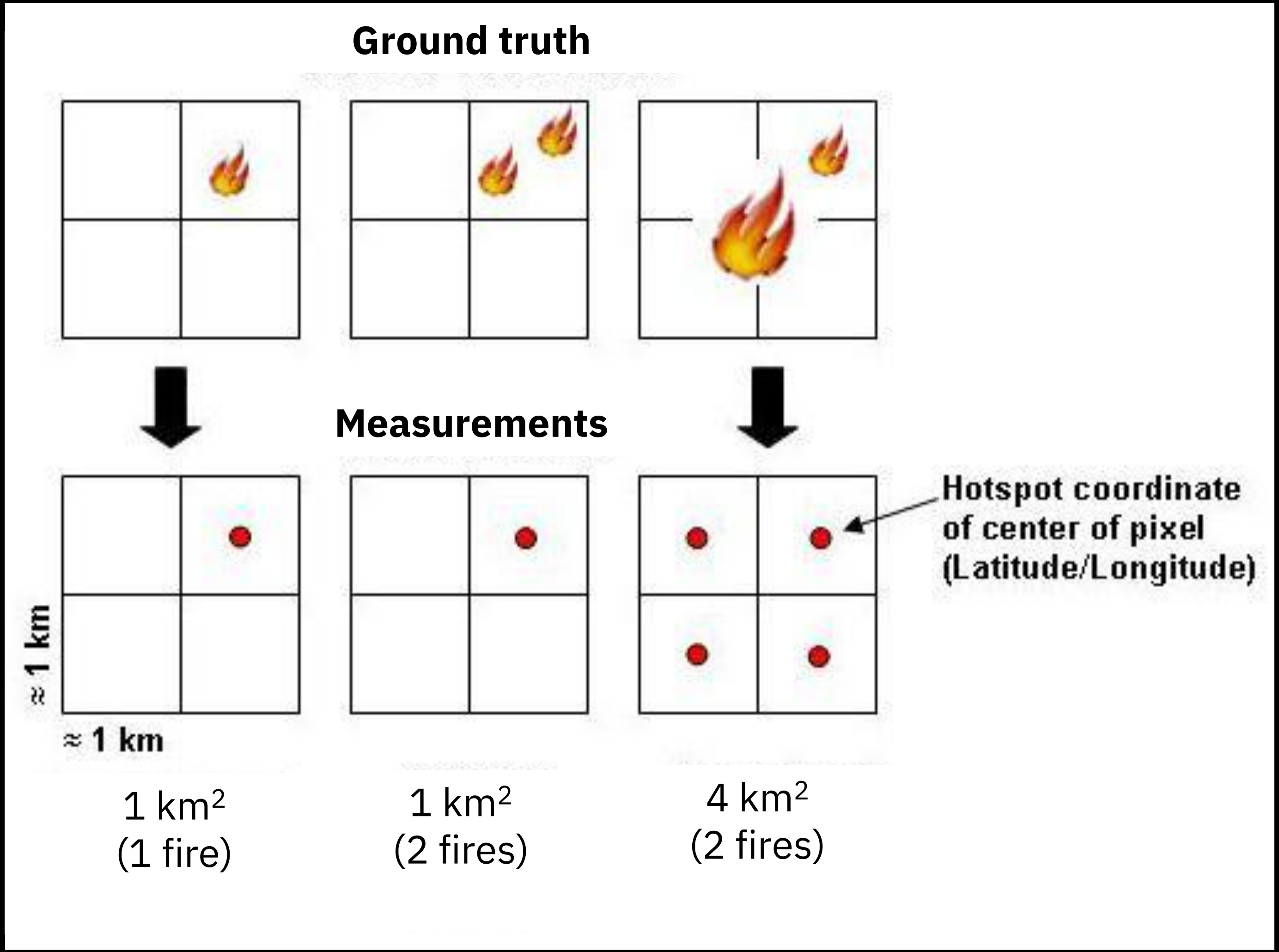
Historical fire data – from hotspots to pixels to estimated areas



1 km^2
(1 fire)

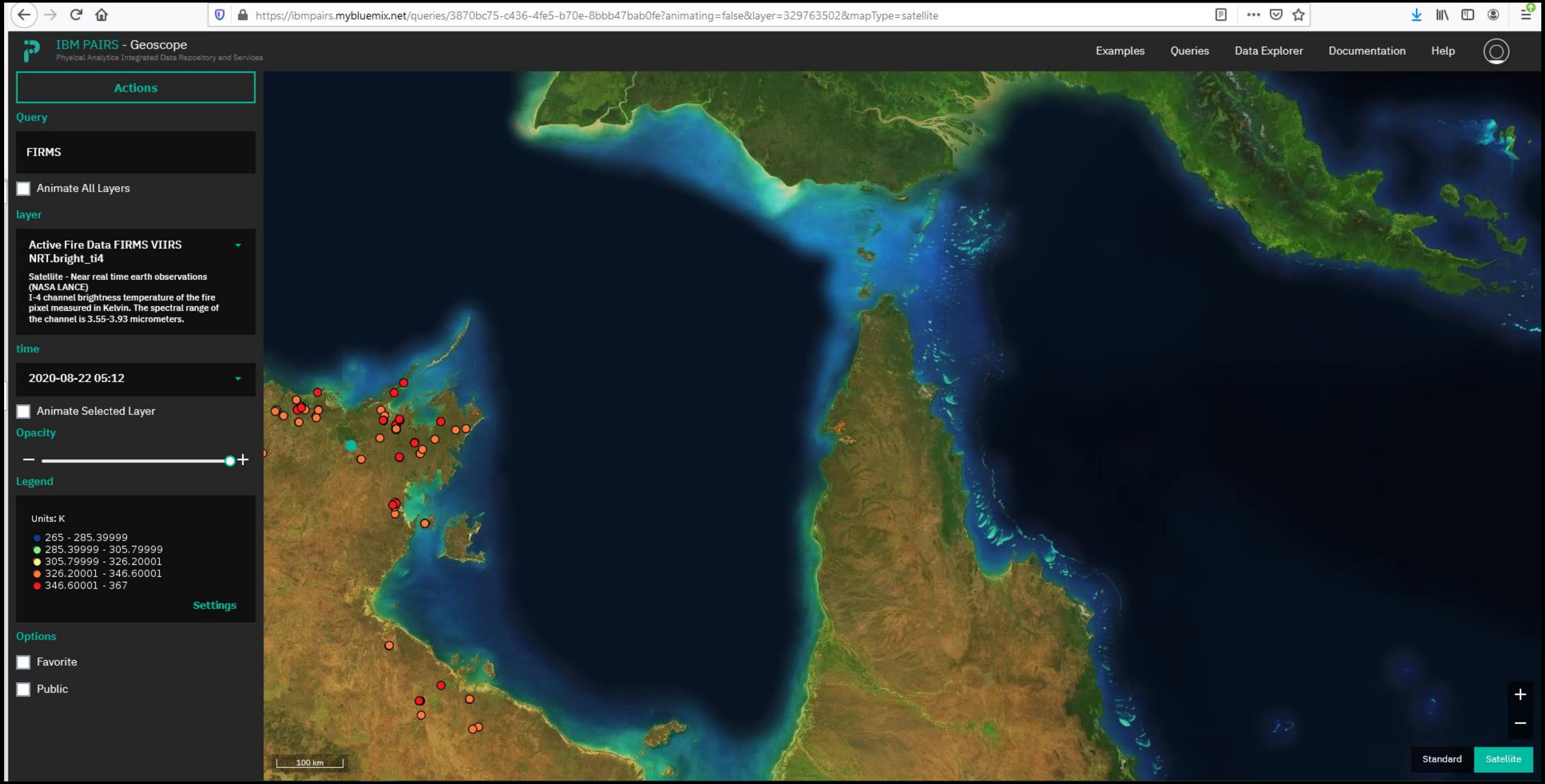
1 km^2
(2 fires)

Historical fire data – from hotspots to pixels to estimated areas



A hotspot plotted using the MODIS thermal anomalies algorithm represents the center of an approximately one-square-kilometer pixel flagged as containing one or more thermal anomalies, which may indicate a fire (upper half of image). The hotspot "location" is the center point of the pixel, which is an approximation of the actual thermal anomaly (lower half of image). Illustration courtesy of NASA FIRMS.

Historical fire data – Raw data



Historical fire data

Basic processing

1. Daily averaged
2. Spatially aggregated over 7 regions
3. Confidence of >75%
4. Inferred hotspot type = 0 meaning a presumed vegetation fire
5. Area estimated by multiplying the along scan pixel size by the along track pixel size.
6. Brightness estimated by averaging the means of both the brightness temperature 21 (obtained from channel 21/22) and brightness temperature 31 (obtained from channel 31).

Columns

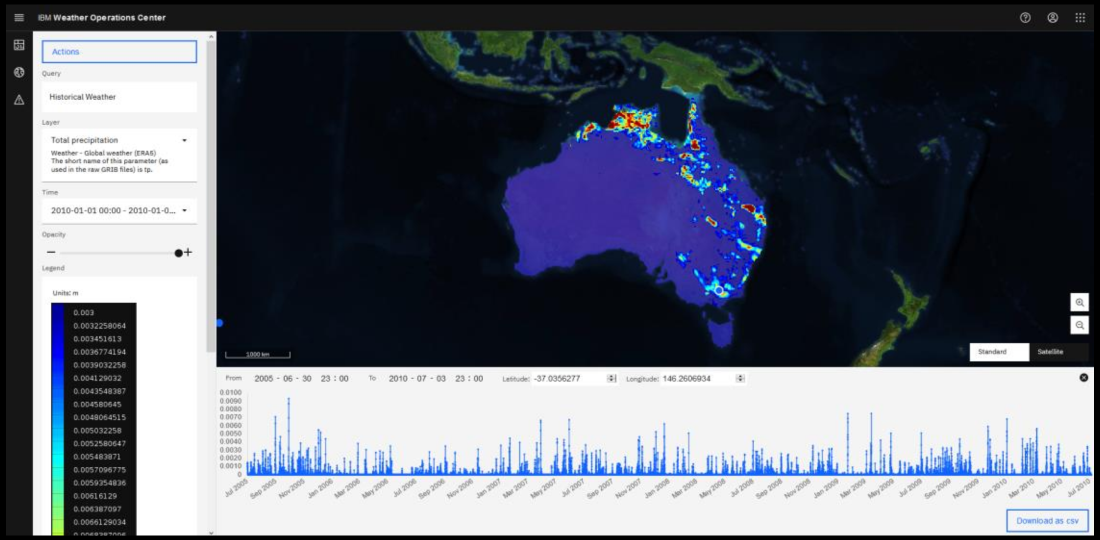
1. Region
2. Date
3. Estimated_fire_area [km^2]
4. Mean_estimated_brightness [K]
5. Mean_estimated_fire_radiative_power [MW]
6. Mean_confidence [%]
7. Std_confidence [%]
8. Var_confidence [%]
9. Count
10. Replaced [Y/N]

© 2018 IBM Corporation

Historical weather data

Basic processing

1. Daily averaged
2. Spatially aggregated over 7 regions



Columns

1. Region
2. Date
3. Parameter
   - ✓ Precipitation [mm/day]
   - ✓ Relative humidity [%]
   - ✓ Soil water content [m3 m3]
   - ✓ Solar radiation [MJ/day]
   - ✓ Temperature [C]
   - ✓ Wind speed [m/s]
4. count() [km2]
5. min(): Minimum of the spatial aggregation
6. max(): Maximum of the spatial aggregation
7. mean(): Average of the spatial aggregation
8. variance: 2nd moment of spatial aggregation

© 2018 IBM Corporation

Historical weather forecast data

Basic processing

1. Daily averaged
2. Spatially aggregated over 7 regions

Columns

1. Region
2. Date
3. Parameter
   - ✓ Precipitation [mm/day]
   - ✓ Relative humidity [%]
   - ✓ Solar radiation [MJ/day]
   - ✓ Temperature [C]
   - ✓ Wind speed [m/s]
4. Lead time [day]
5. count() [km^2]
6. min(): Minimum of the spatial aggregation
7. max(): Maximum of the spatial aggregation
8. mean(): Average of the spatial aggregation
9. variance: 2nd moment of spatial aggregation

© 2018 IBM Corporation

Land class data

Basic processing

1. Spatially aggregated over 7 regions
2. Data is normalized to 100 [%]

Columns

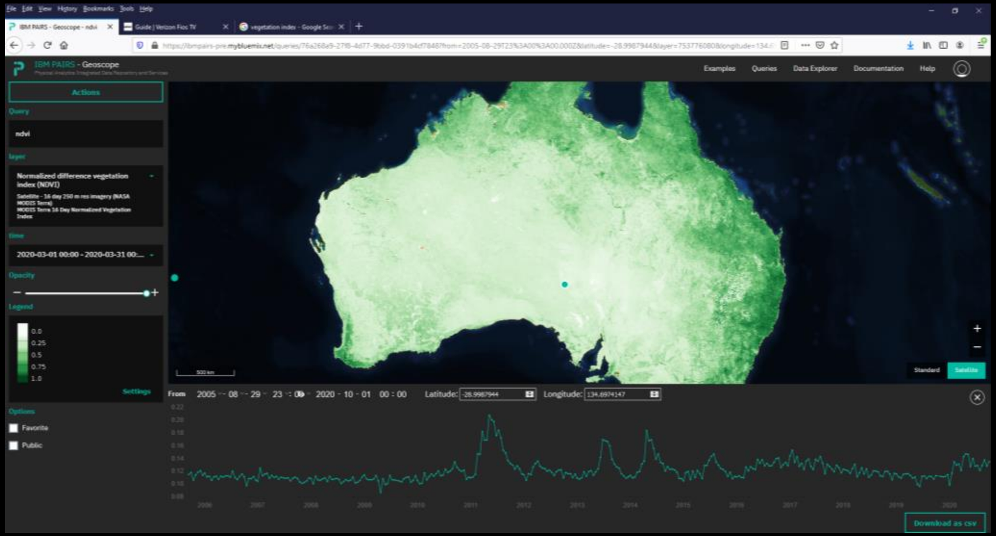
1. Region
2. Type
   - ✓ Shrubs [%]
   - ✓ Herbaceous vegetation [%]
   - ✓ Cultivated and managed vegetation/agriculture (crop
   - ✓ Urban / built up [%]
   - ✓ Bare / sparse vegetation [%]
   - ✓ Permanent water bodies [%]
   - ✓ Herbaceous wetland [%]
   - ✓ Closed forest, evergreen, broad leaf [%]
   - ✓ Closed forest, deciduous broad leaf [%]
   - ✓ Closed forest, unknown [%]
   - ✓ Open forest, evergreen broad leaf [%]
   - ✓ Open forest, deciduous broad leaf [%]
   - ✓ Open forest, unknown definitions [%]
   - ✓ Open sea [%]

© 2018 IBM Corporation

Vegetation index data

Basic processing

1. Monthly aggregated
2. Spatially aggregated over 7 regions



Columns

1. Region
2. Date
3. Vegetation_index_mean
4. Vegetation_index_max
5. Vegetation_index_min
6. Vegetation_index_variance

© 2018 IBM Corporation

Crowdcast April 5, 2021 – Predict Wildfires in Australia

- Welcome and Call for Code Overview (Upkar)
- Overview & Summary Results (Susan)
- The datasets (Hendrik)
- A technical overview of the models from 3 teams
  - yau_yee_Italy Team (Marco)
  - Data Warriors (Albert)
  - NA Team (Brianna, Ned)
- A summary comparison of the modeling approaches (Wiktor)
- Students using the challenge as part of their university program
  - Dr Manar, Emad, Abdelhady
- What's next (Susan)
- Call for Code Close and Q&A (Upkar)

Call for Code Spot Challenge for Wildfires

The yau_yee_Italy’s Approach for developing the Solution

Team Members

Marco De Ieso

Andrea Ongaro

Giulia Plotti

Guglielmo Sanchini

22 Febbraio 2021

The IBM Wildfire Challenge

Time-series problem based on daily data provided by **Earthdata Nasa** and **Pairs Geoscope** describing wildfire extension per region in Australia, meteo condition and other properties

Five **dataset** provided:
Historical Wildfire
Historical weather
Historical weather forecasts
Historical vegetation index
Land classes

Predict the size of the fire area in km squared by region in Australia for each day in **February 2021** using data available up to January 29th

The yau_yee_Italy Team

Marco De Ieso
IBM Senior Data Scientist – Leader

Andrea Ongaro
IBM Data Scientist

Giulia Plotti
IBM Data Scientist

Guglielmo Sanchini
IBM Data Scientist

© 2021 International Business Machines Corporation

The phases of the project

October	November	Decembe r	January	February

Preparation → **Solution Development & Submission** → **Review & Dissemination**

Preparation

Study of the Literature and approaches already implemented

Development

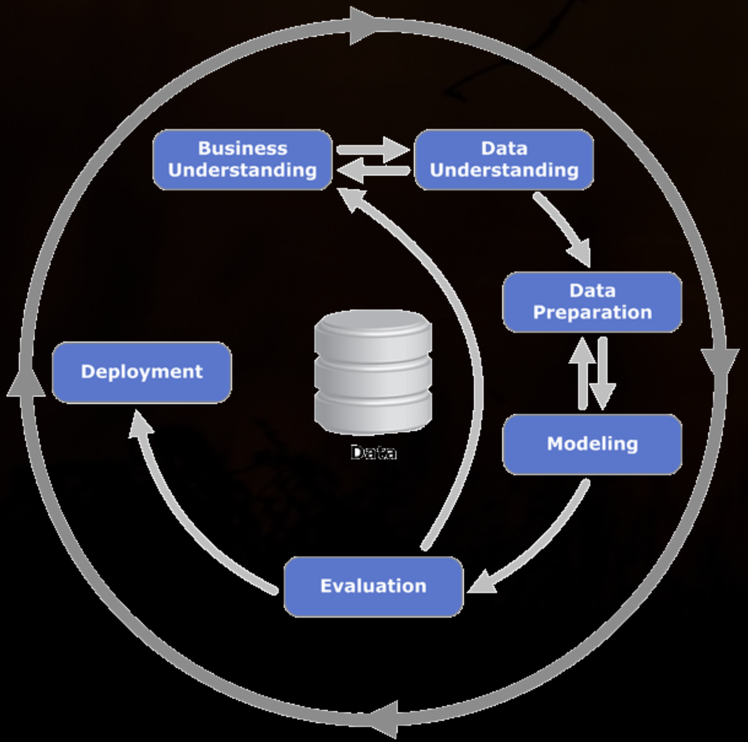
Application of the CRISP-DM Methodology to the wildfire challenge

Review & Dissemination

Sharing of results, feedbacks and lessons learned

CRISP-DM: Overview of the Methodology

Cross-industry standard process for data mining (CRISP-DM) is an open standard process model that describes common approaches used by data mining experts. It includes six main phases:



Business Understanding	Understand business Context, Objectives, and Requirements
Data Understanding	Data collection, review and exploration of available data
Data Preparation	Construction of the data set(s) for modeling
Modeling	Application of algorithms and procedures
Evaluation	Model performance Evaluation & Business Value Estimation
Deployment	Result application/Operational implementation

Data understanding – Variables that impact wildfires

Territory

Regions

Number and intensity of fires vary across different regions
Events in neighbouring regions can influence wildfires in a given territory
Autoregressive component (the presence of wildfires in the preceding days)

Seasonality

Wildfires are particularly intense during the "Bushfire Season" which goes from October to February.

This observation will affect the way in which training and test set are divided.

Soil and Atmospheric conditions

Weather & soil

Precipitations and Drought is strictly related with fires
Comparison between historical and historical forecasted weather

Vegetation

There is a positive correlation between the variation in the vegetation index and the intensity of fires.

Land usage

Land utilization might be relevant for predicting the extension of wildfires. However, this data was available only as a constant – for every region; thus it was ***not*** *included in the model.*

Data Preparation

Perimeter of analysis

Strategy: one model for each couple (T,R) where T is the time of advance and R is a Region

According to the strategy we built for every (T,R) the subset of data with which we want to feed into the model.

Computation of the target variable – what we want to predict – for each time unit T included in the perimeter.

Preprocessing

We created all the covariates that might be relevant for predicting the target value, based on the insights from the Data Understanding phase.

Before computing variables, it is necessary to assess the quality of data, by eliminating outliers, correcting any erroneous data or replacing missing one.

Partitioning Strategy

The mining table is divided in different subset to assure that the model created is robust and precise :

1. **Training set**
2. **Test set**

We used the last 3 February and January 2021 as test set. For each test month, we trained the algorithm until month – 1.

Data Preparation

Preprocessing: Feature Extraction

We created four classes of variables. In order to reduce the complexity of the problem and to avoid multicollinearity we applied PCA technique

AUTOREGRESSIVE VARIABLES

Lag features of the target variable
Rolling functions [min, max, mean, sdev] applied to the target variable
Distance in days from the last fire event

WEATHER DATA

Using both historical data and forecasting we extracted:
Weather variables(precipitation, temperature,...)
Lag features of weather variables.
Index on drought like PET and SPI

SEASONALITY

The Australian Bureau of Meteorology defines five “fire danger season”, being times of peak bushfire activity. Therefore, we add the following features:
Month
Day in the year

VEGETATION INDEX

Derive daily VI by interpolation. Using this, we then construct:
Rolling mean
Lag features

Modeling

Choice of the Algorithms

The main criteria for selecting the algorithms has been the ability to perform well on out-of-sample

Supervised Algorithms
- *XGboost*
- *Random Forest*
- *LightGBM*

Feature Selection

Several methods have been applied to select the set of relevant features

Methods adopted
- *Correlation Matrix*
- *Backward Elimination*
- *Recursive Feature Elimination*

Tuning of Parameters

Optimization of those parameters whose values cannot be estimated from the data has been performed

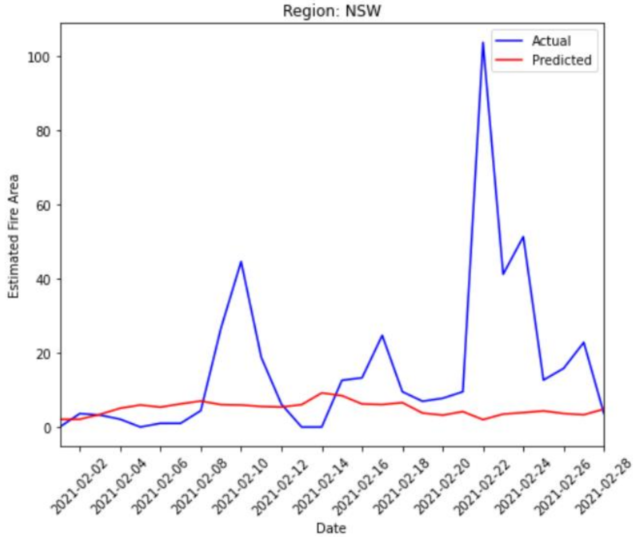
Techniques experimented:
- *Grid Search*

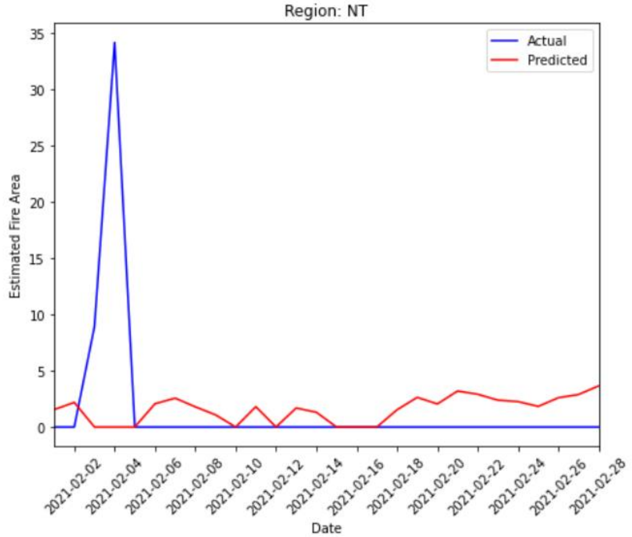
Ensembling

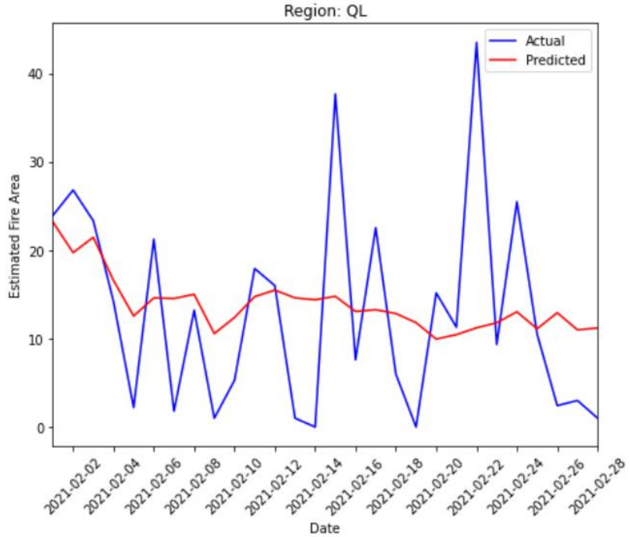
The accuracy is increased by combining the algorithms to construct better predictive performance

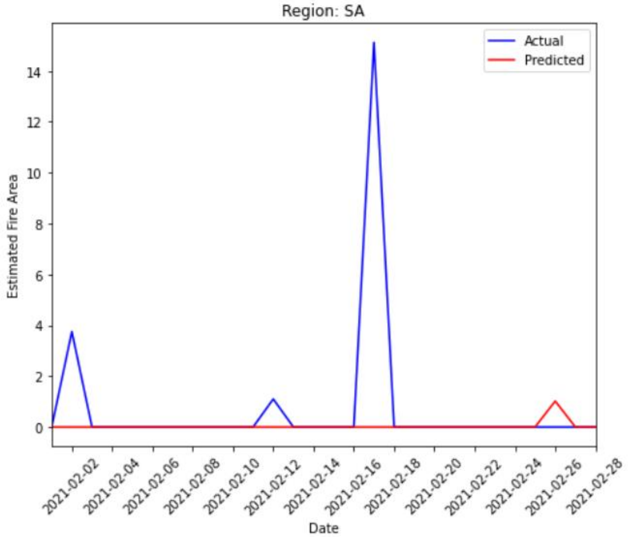
Ensembling Techniques
- Averaging
- Based on the forecast advance

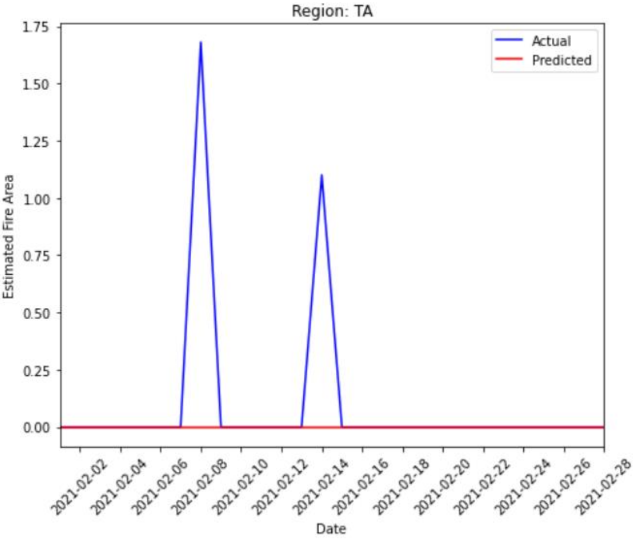
Results
Actual vs Predicted on Feb 2021

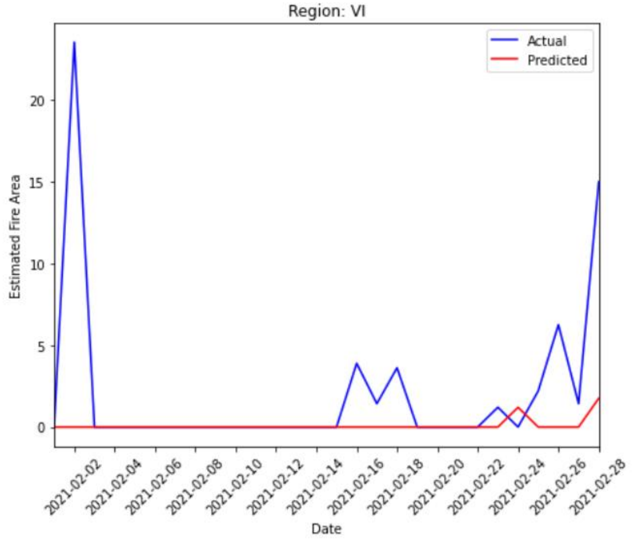


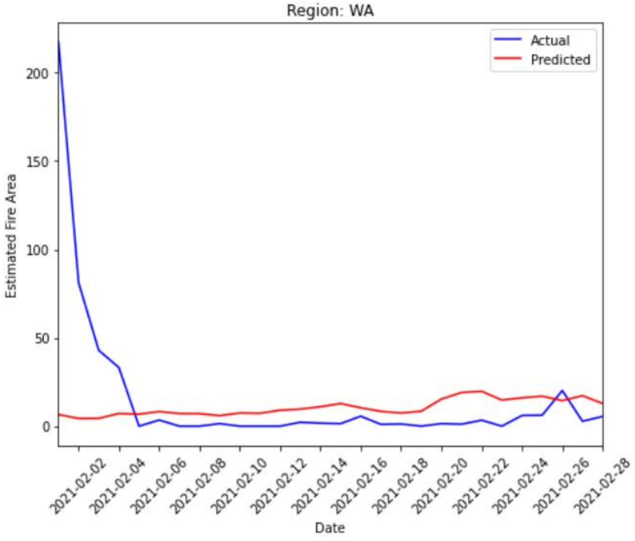






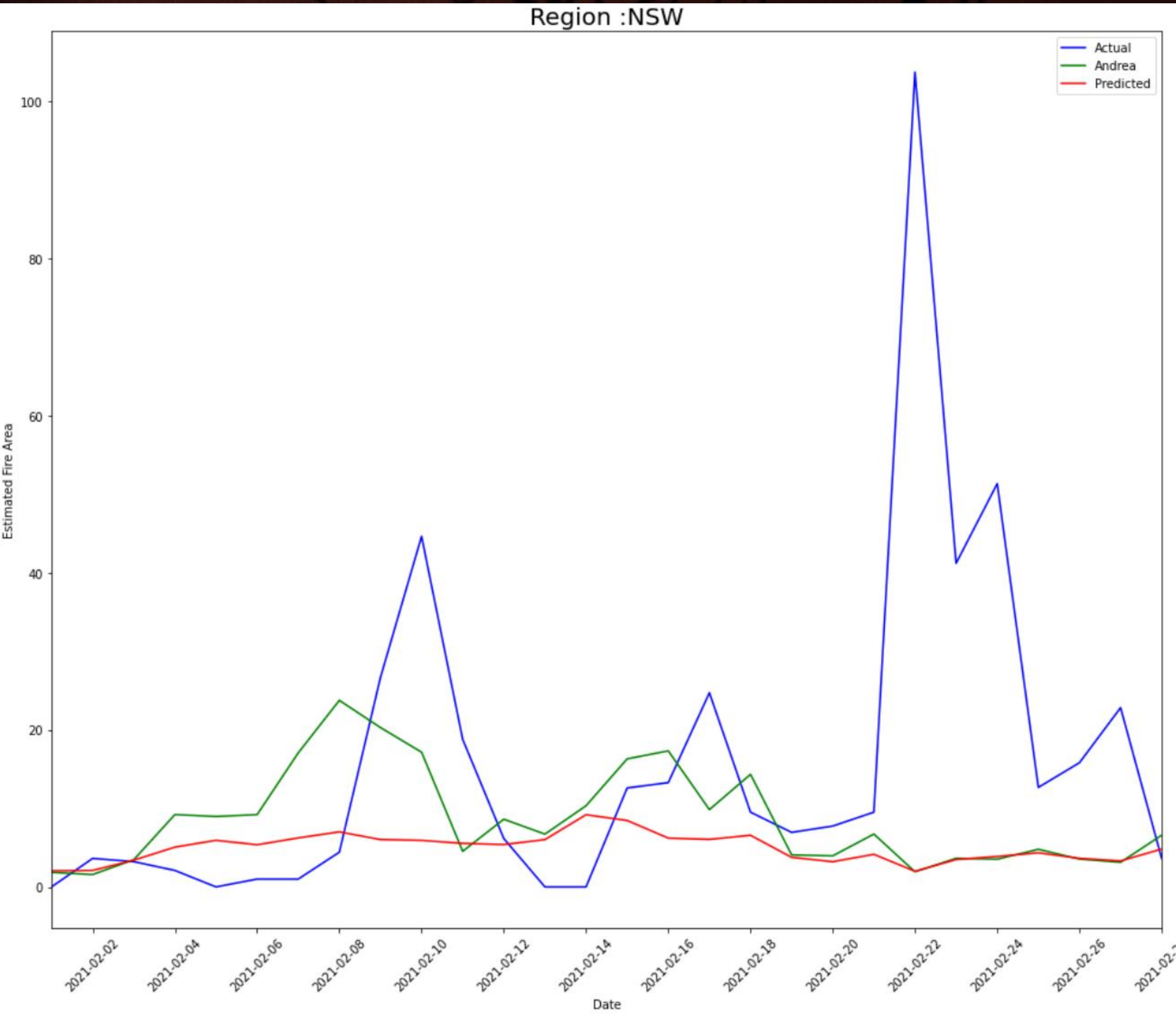


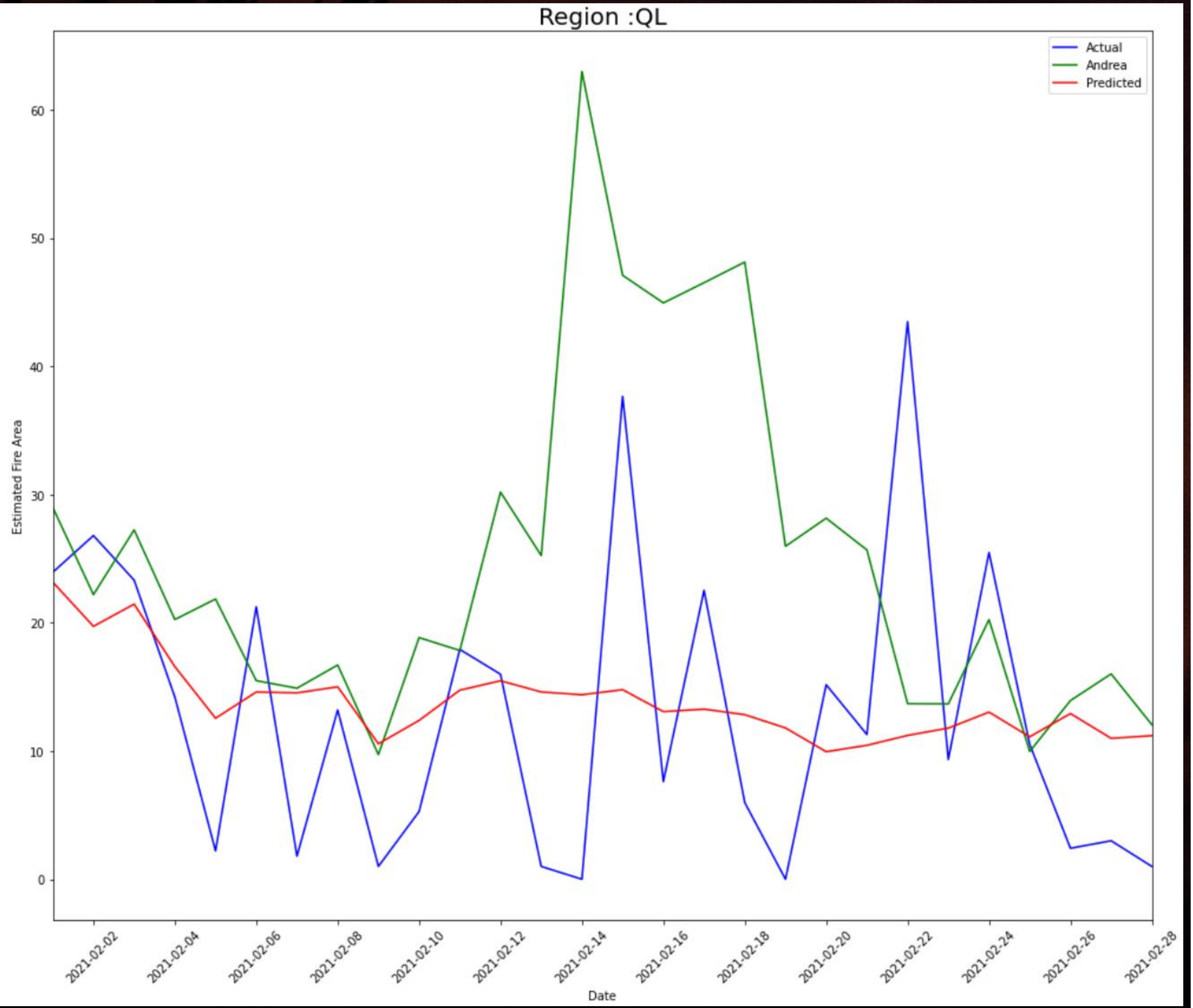




RMSE: 19.96
MAE:6.94
TOT:9.54

Results
Actual vs Predicted on Feb 2021





Call for Code Spot Challenge: Australian Wildfires

Data Warriors

Data Warriors: Introduction

Albert Um

- Data Scientist
- Bachelors in Finance
- Manufacturing Consultant
- Certified in Production and Inventory Management

Shruti Chaturvedi

- Azure AI Engineer
- DevOps Developer at Connexa.ai
- Harvard WECode'21 Lead Tech Fellow
- Bit Project ML Instructor

Divyansh Choubisa

- Machine Learning Engineer
- Bachelors in Computer Science Engineering

Competition Objective

Predict daily fire areas(in km2) by region for February 2021:

- NSW - New South Wales
- NT - Northern Territory
- QL - Queensland
- SA - South Australia
- TA - Tasmania
- VI - Victoria
- WA - Western Australia

Leaderboard Evaluations:

➢ 80% MAE (Mean Absolute Error)
➢ 20% RMSE (Root Mean Squared Error)

Process Outline

Data Study

MCD14DL(wildfires)
ERAS data(weather)
MOD13Q1(vegetation)

Preprocessing

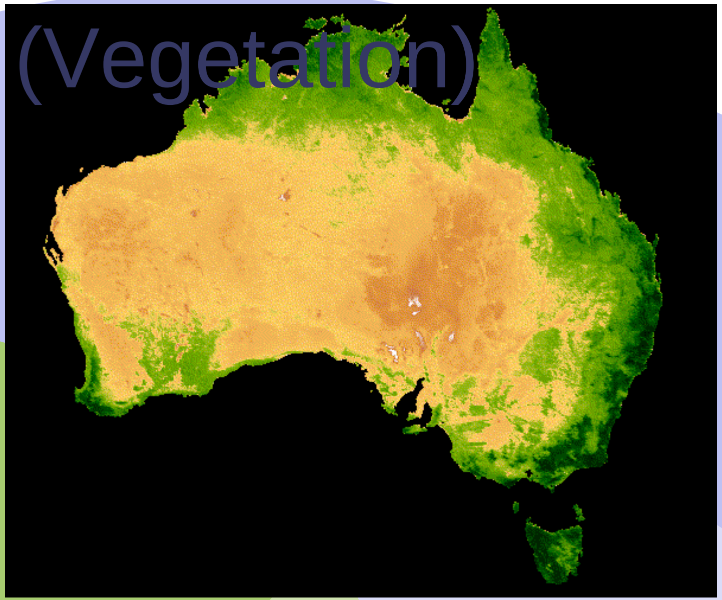
Feature Engineering
Windowed Datasets

Model Evaluation

Dilated Convolution
Neural Network

Data Study

- MCD14DL (Wildfires)
- ERAS Data (Weather)
- MOD13Q1 (Vegetation)



Historical_Wildfires.csv

HistoricalWeather.csv

VegetationIndex.csv

Preprocessing Features

Weather & Vegetation Means

Precipitation, Relative Humidity, Soil Water Content, Solar Radiation, Temperature, Wind Speed, NDVI

Log Scale

Wildfires have a power law-like distribution. Log transform the estimated fire areas.

Surface Area

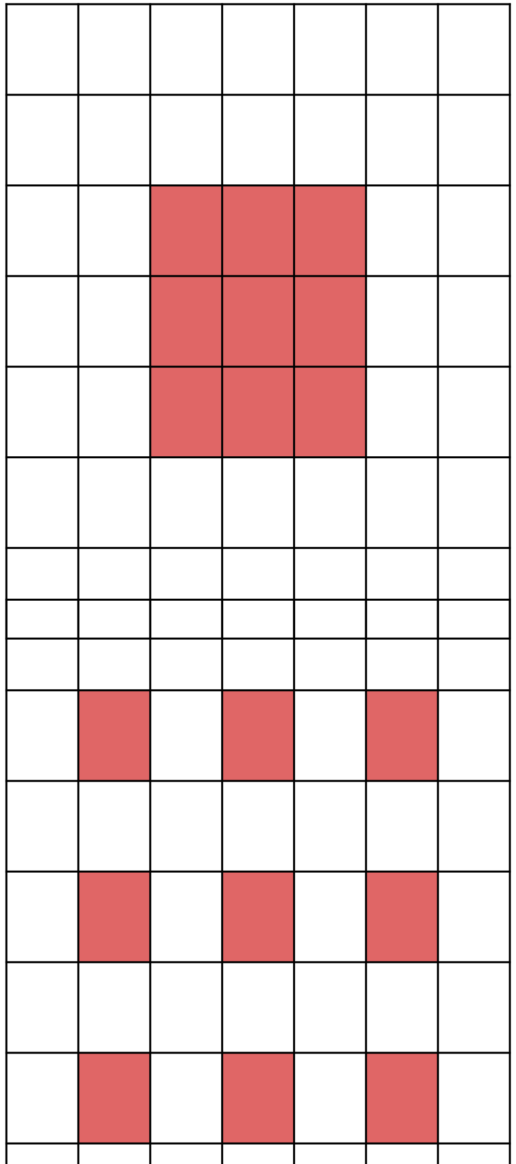
Feature engineer the expansion variable for wildfire areas.

Feature Engineering: Surface Area

Let :
SA = Surface Area
C = Count

Conglomerated

Assume the fire area pixels are united to one large block per region.



$$SA_c = 4 * \sqrt{Area}$$
$$SA_c = \sqrt{Area}$$

Separated

Assume the fire area pixels are dilated and each pixel has equal area.

$$SA_s = \begin{cases} C * \sqrt{\frac{Area}{C}} & \text{if } C > 0. \\ 0 & \text{otherwise.} \end{cases}$$

Model Evaluation

Predict up to February 28, 2021.

Dataset ends on January 18, 2021.
The output steps will be 41 days
(41 days + Jan. 18 = Feb 28)

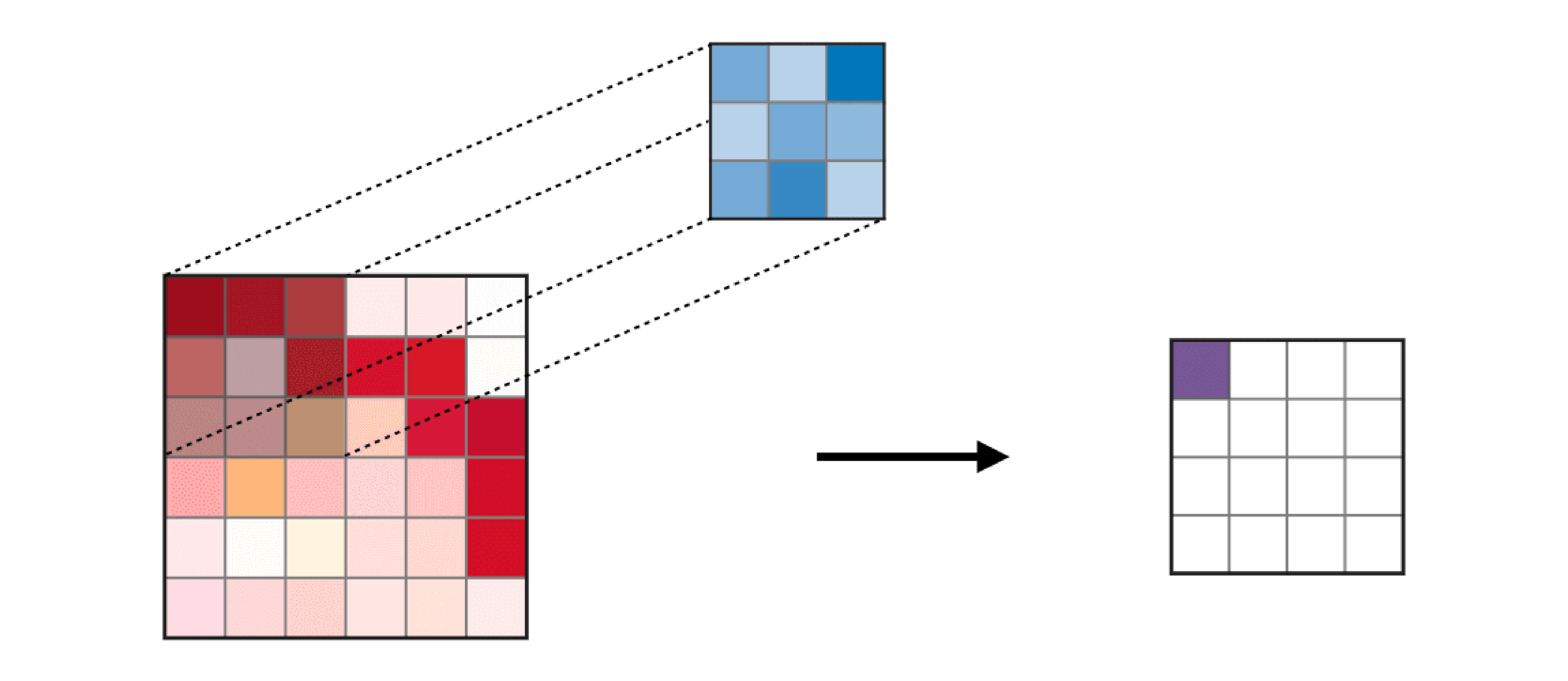
Training set

January 01, 2005 -
November 30, 2020

Testing set

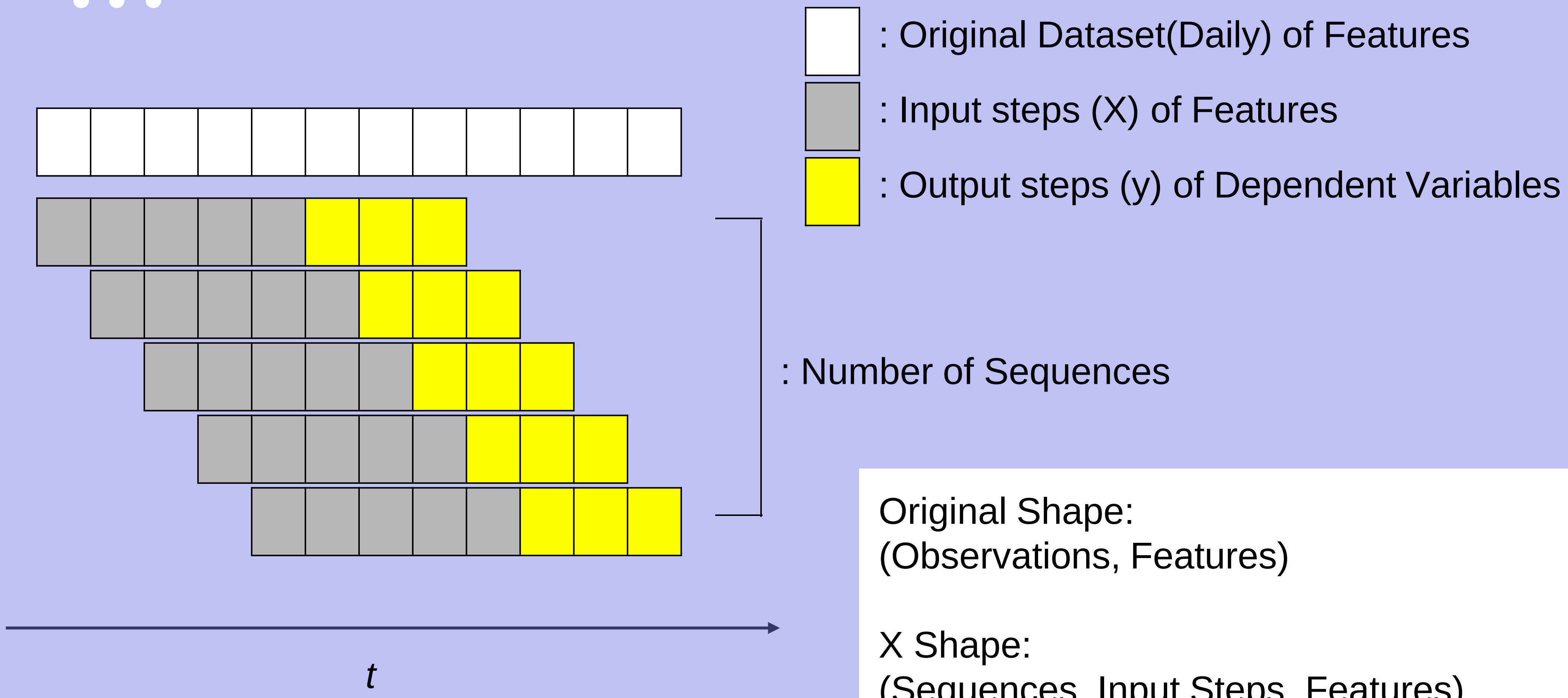
December 01, 2020 -
January 11, 2021

Convolutional Neural Network



Source: https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks

Windowing Dataset



Original Shape:
(Observations, Features)

X Shape:
(Sequences, Input Steps, Features)

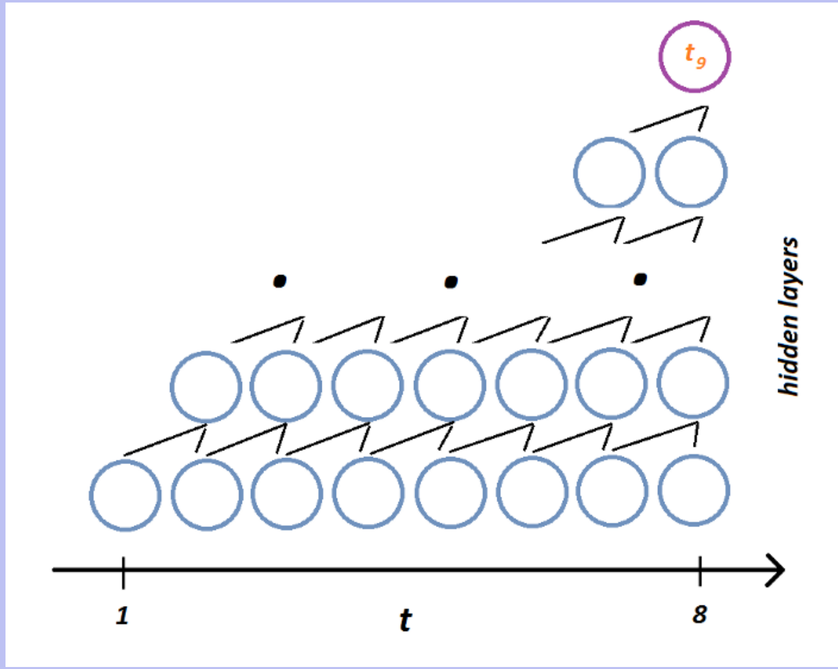
Y Shape:
(Sequences, Output Steps, Dependents)

Conv1D Layers

Input Steps: 120, Output Steps: 41

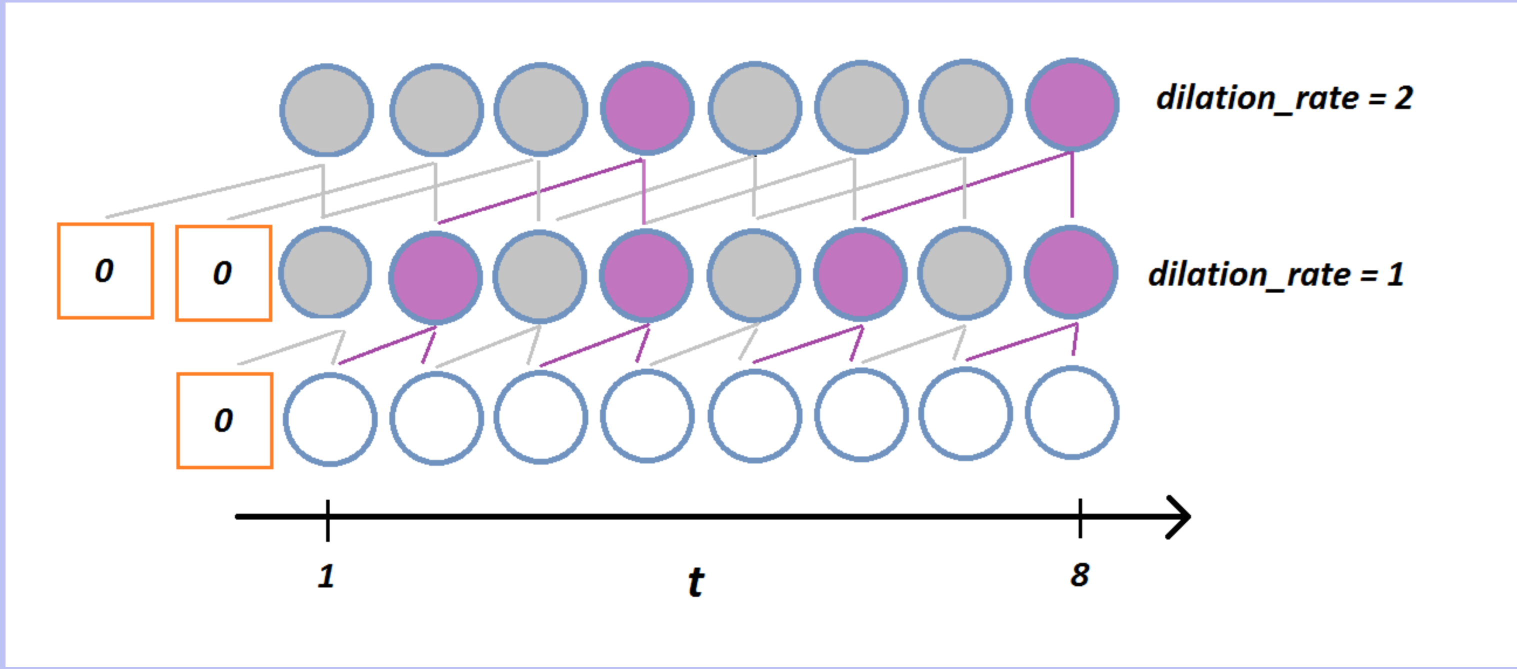
Without Padding/Dilation

- • •



With Padding and Dilation

- • •



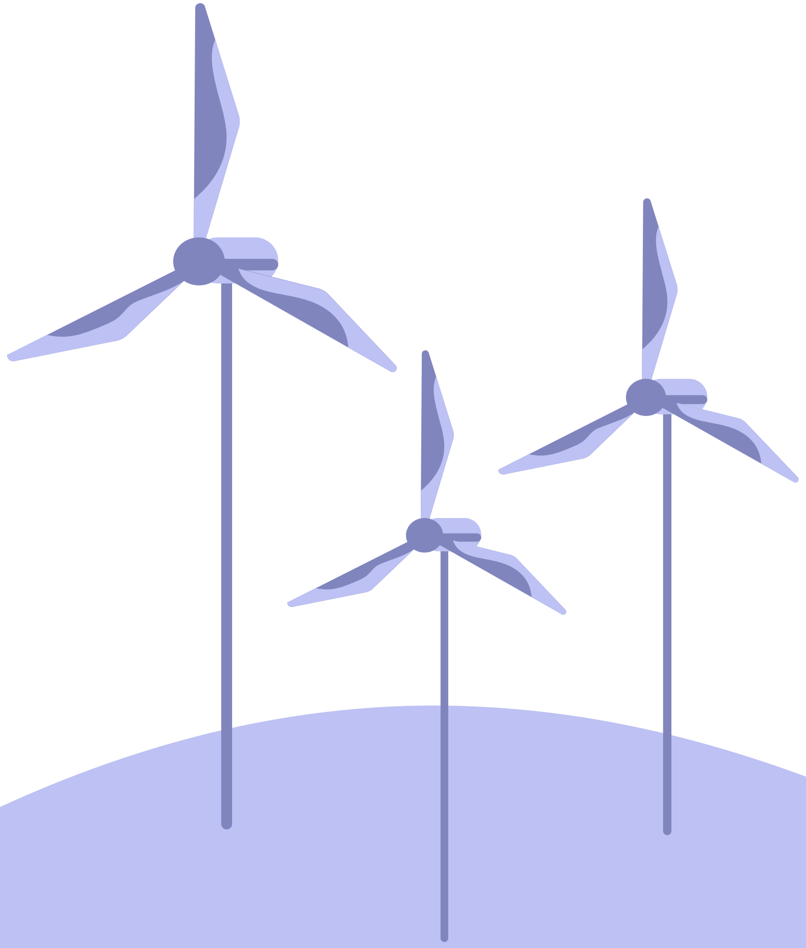
Current Standings

February 01, 2021 - February 28, 2021

Mean Absolute Error: 7.03

Root Mean Squared Error: 19.60

Further Steps



AutoRegressive

Incorporate autoregressive mode into model.

Google Earth Engine

Reaggregate data to administrative level 2 or custom hexagon shape.

Thank you!

https://github.com/albertum1/cfc_team

Australian Wildfires Spot Challenge

Team NA

Introduction:

Brianne Boldrin

Business Analyst & Technical Delivery Manager

Goals for the challenge:

- Increase experience with Python
- Exposure to new real data sets & use cases
- Practice!

Ned Bader

Advisory Engineer & Master Inventor

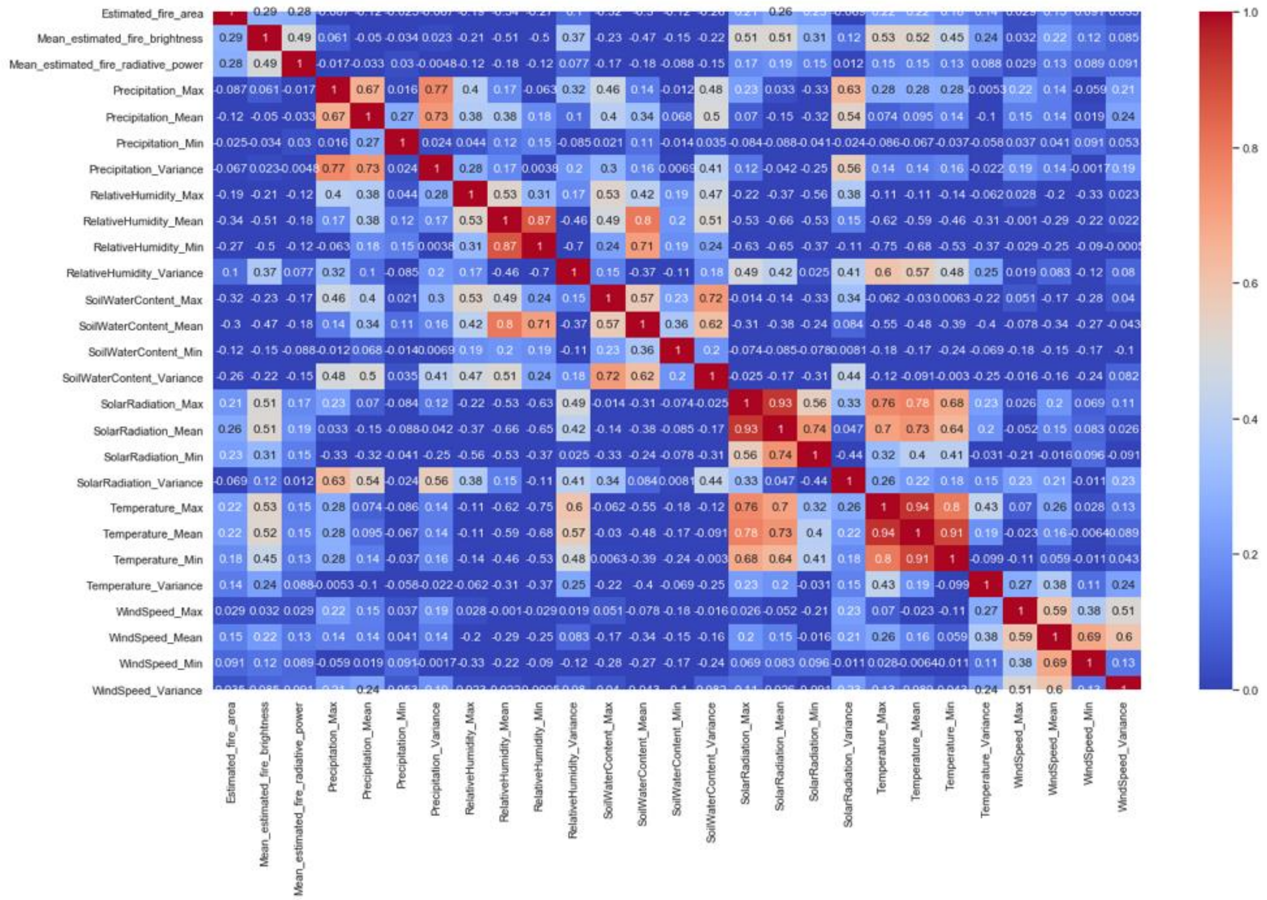
This data challenge allowed me to grow in:

- Understanding real problems
- Working with real data sets
- Evaluating different processing methodologies

Data Analysis

Data Analysis:

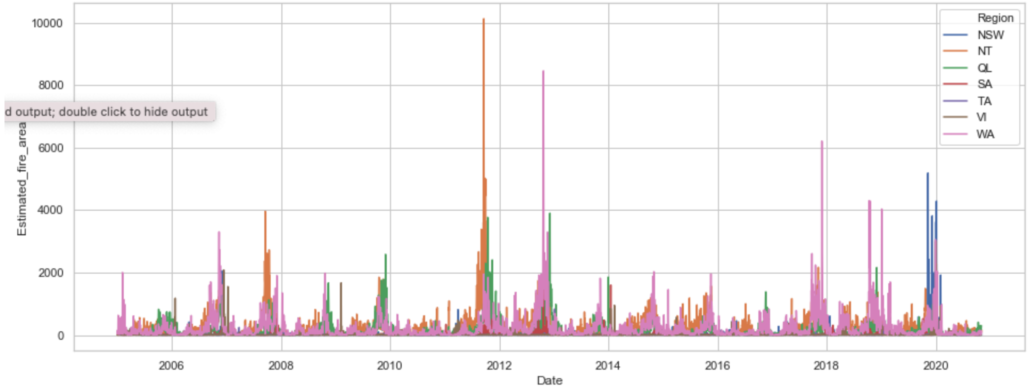
Correlation Analysis: Identifying Similar Variables



Variable Statistics:

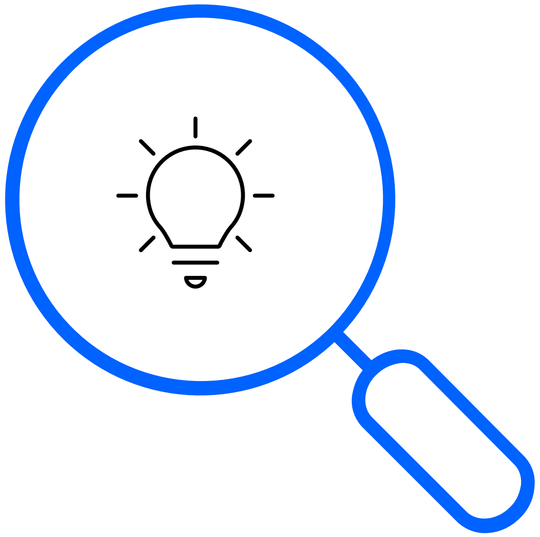
	count	mean	std	min	25%	50%	75%	max
Estimated_fire_area	26406.0	144.297966	314.453847	1.000000	8.911875	3.843409e+01	146.951278	10120.943170
Mean_estimated_fire_brightness	26406.0	319.662078	8.862005	290.700000	313.933333	3.197844e+02	325.403144	381.950000
Mean_estimated_fire_radiative_power	26406.0	83.621258	67.510022	0.000000	44.150391	6.713333e+01	103.123611	2178.600000
Mean_confidence	26406.0	87.574735	4.371972	76.000000	85.000000	8.777143e+01	90.498403	100.000000

Actual Fire Area by Region



Hypotheses:

- Predicting data for entire month will be challenging, as daily averages are heavily correlated to each other
  - Strategy: Find the average fire area weekly vs day by day to account for variability
- Yearly weather conditions may predict the type of wildfire year
  - Strategy: look for an annual pattern for January wildfires
- If we know the weather conditions, we can better predict the wildfire area.
  - Strategy: Use the weather conditions to predict wildfire area, eliminate weather conditions that are correlated to each other to reduce the number of variables.



Methodology

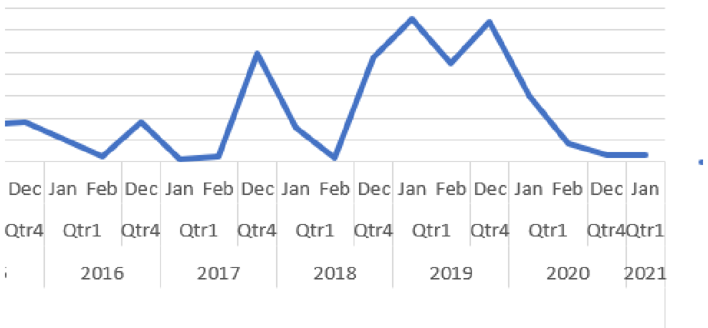
Method Pivot and Smash:

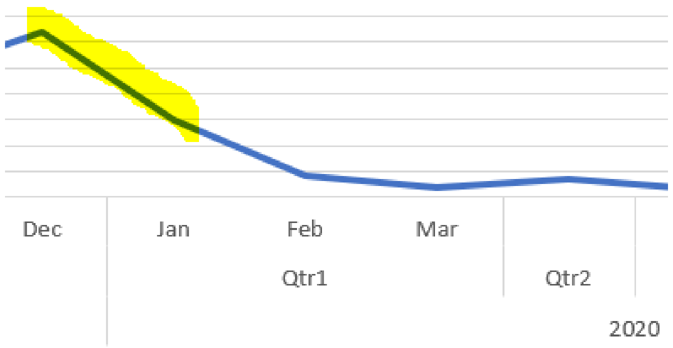
Using Pivot functionality from EXCEL

- Initially used to provide visualization of the fire data and periodization across the years
- Simple predications were created to have a submittal for the first phase of the competition

Final Process (Trend analysis) included the following steps

- Pivoting the historical fire data
- Normalized the daily averages to account for zero dates (e.g. data cleaning)
- For each region
  - Checked trends (ratios) for December and January for the last 5 years
  - Determined which trend fit the present December/January data
  - Then used the ratio of the best fit to calculate this year's Feb weekly averages





Avg					
	Dec	Jan	Feb	Feb/Jan	Jan/Dec
2016	178.77	94.97	20.27	0.213426	0.531224
2017	172.58	6.58	21.87	3.324724	0.038114
2018	493.81	153.32	18.84	0.122892	0.310477
2019	472.03	651.16	446.92	0.686345	1.379493
2020	634.84	289.09	76.71	0.265354	0.455371
2021	27.99	31.73			1.133789

Retrospective

What we learned:

- Start by understanding the data & problem
  - What values are missing? What variables may be correlated? How does each variable impact our fire prediction? What is the daily fire variability?
- Important steps for cleaning data sets
  - Missing values, etc.
- Start simple - don't get trapped in micro-calculations
- Power of a diverse team
  - With teammates from different backgrounds, sharing results and ideas helped our models!

If we had more time:

- More automation - data collection, cleaning, and processing
- Time series analysis
- Including more data (geospatial data, etc.) and external data sources --- if area burns one year, would it be likely to burn again?

Thank you

Call for Code Spot Challenge for Wildfires Predictions – A summary comparison of the modeling approaches

April 5, 2021

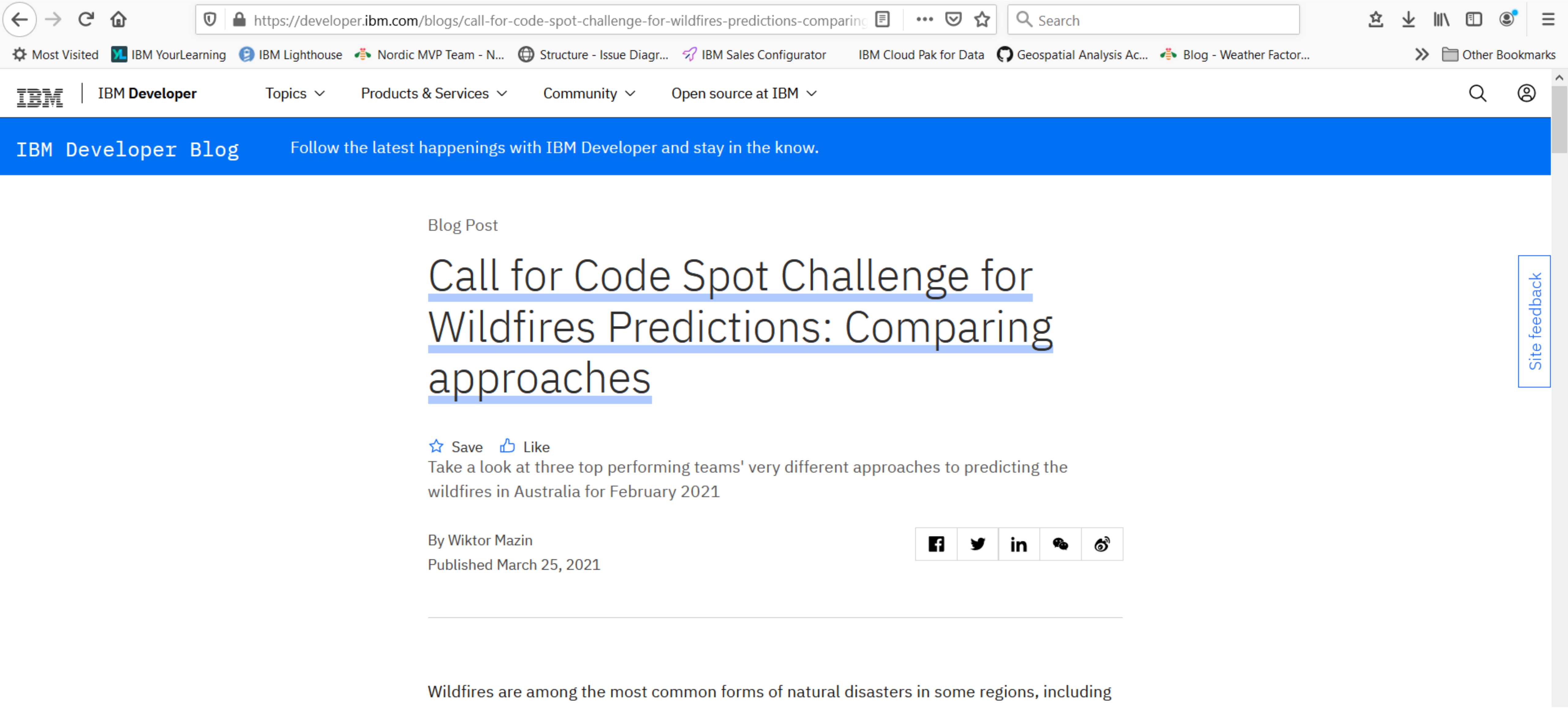
Introduction:

Wiktor Mazin
Chief Data Scientist, IBM

Data & AI, MVP team, Nordics
wiktor.mazin@ibm.com

© 2018 International Business Machines Corporation

Comparing approaches



https://developer.ibm.com/blogs/call-for-code-spot-challenge-for-wildfires-predictions-comparing-approaches/

© 2018 International Business Machines Corporation

Crowdcast April 5, 2021 – Predict Wildfires in Australia

- Welcome and Call for Code Overview (Upkar)
- Overview & Summary Results (Susan)
- The datasets (Hendrik)
- A technical overview of the models from 3 teams
  - yau_yee_Italy Team (Marco)
  - Data Warriors (Albert)
  - NA Team (Brianna, Ned)
- A summary comparison of the modeling approaches (Wiktor)
- Students using the challenge as part of their university program
  - Dr Manar, Emad, Abdelhady
- What's next (Susan)
- Call for Code Close and Q&A (Upkar)

ALPHAWARE

Mohamed Moataz
Amna Ibrahim
Afreen Saif
Ibrahim Maher
Mohamed Abdelhady

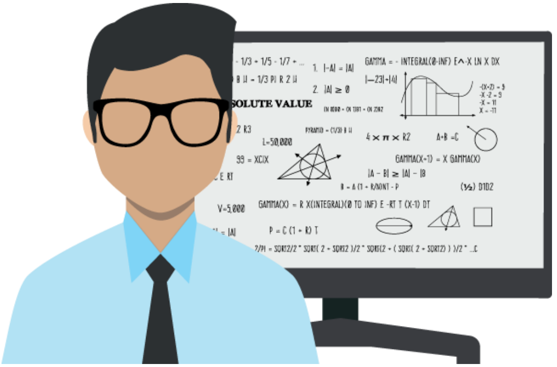
Team in Sharjah University Software Engineering Class

ROLES

SOFTWARE ENGINEER

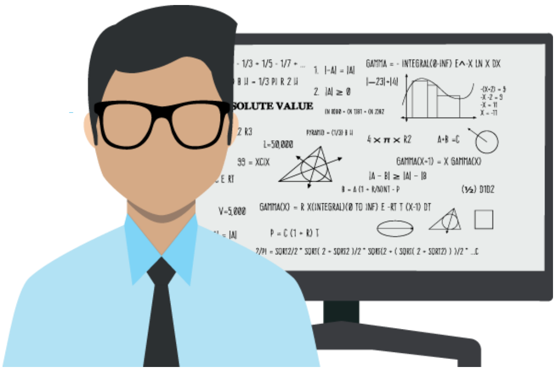
- Modeling Database
- Implementing Backend APIs
- Implementing Frontend Views
- Networking Protocol
- Authentication and Authorization

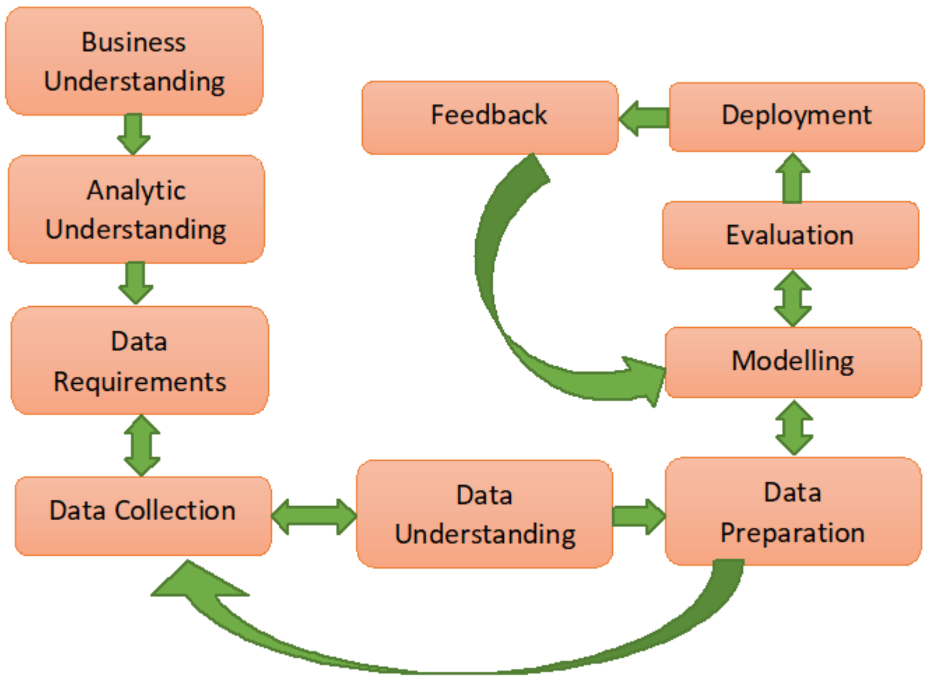
DATA SCIENTIST



- Follow Data Science Methodology
- Studying Data Requirements
- Collecting Data
- Feature Engineering and Normalization
- Building Prediction Model
- Maintaining and Upgrading Prediction Model

DATA SCIENTIST





- Programming Language: Python
- Machine Learning Technology: PyTorch
- Visualization: PyPlot
- Mathematical Computation: NumPy

SOFTWARE ENGINEER

TECHNOLOGY STACK

- Database: MySQL
- Frontend: Bootstrap 4 and React
- Backend: Python Framework (Django or Flask)
- Networking: TCP/IP and REST API

FEATURES

- Pinpointing Nearest Shelter
- Displaying Charts
- Send Warning Messages (SMS)
- Raising Awareness Through Social Media Platforms

4SIGHT

PREDICTION FOR PREVENTION

Wildfire Prediction in Australia

Emad Aldawoud (U18102200@sharjah.ac.ae)

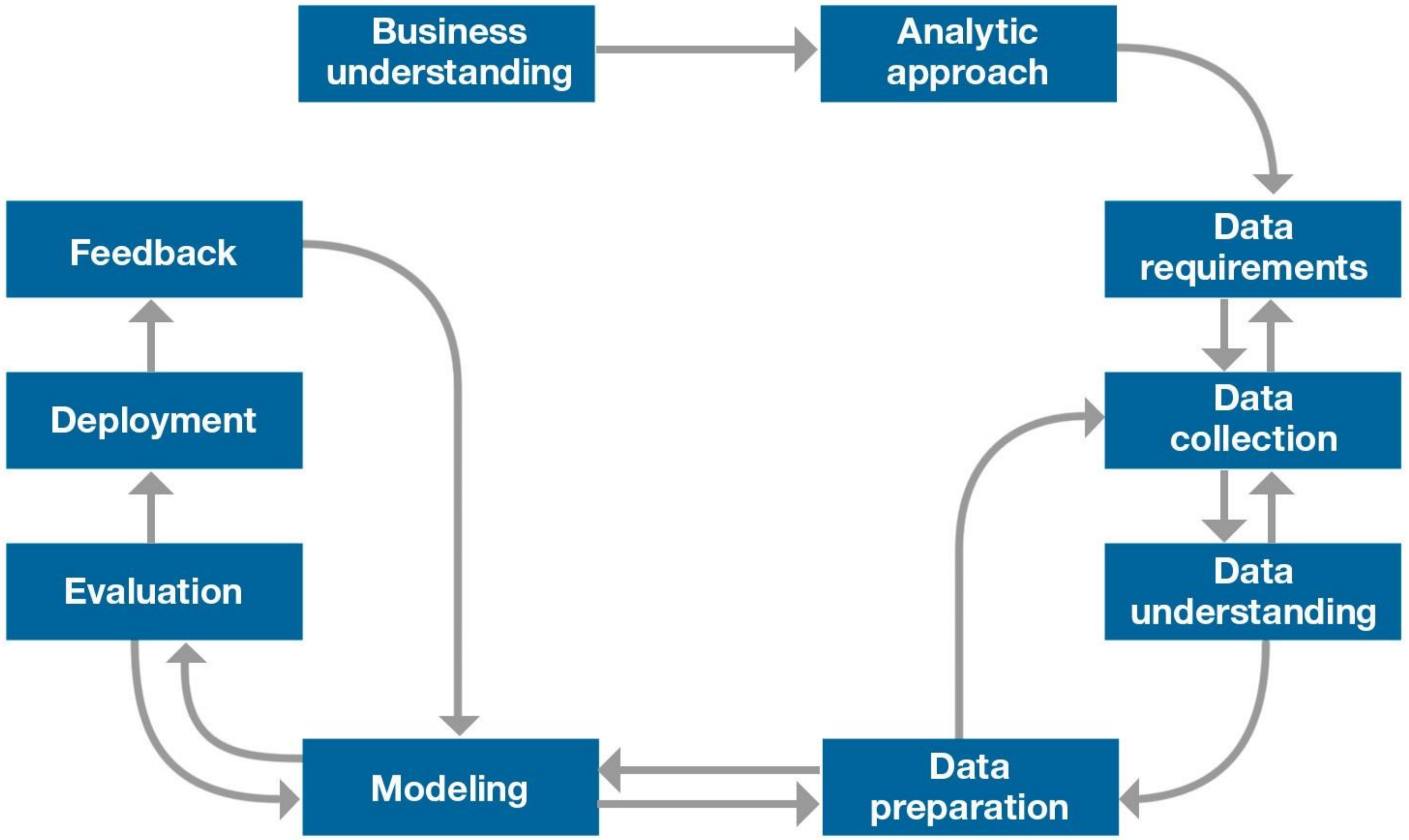
Mohamad Abdallah

Omar Ibrahim

Amro Tariq

Rashad Zingstra

Data Science Methodology



References for Data

https://github.com/Call-for-Code/Spot-Challenge-Wildfires

Fires in Australia: https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data

Weather data:
https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset

Forecast data: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs

Final Processed Data

Input

Target

Date	Region	Season	TemperatureMean	PrecipitationMean	RelativeHumidityMean	SoilWaterContentMean	SolarRadiationMean	WindSpeedMean	Estimated_fire_area
3/1/2001	NSW	Autumn	23.8109531	0.0363	0.528446777	130.313305	8.820266699	2.695024485	20.06
3/2/2001	NSW	Autumn	23.2040798	0.072	0.500738441	130.631899	9.009442656	3.203867011	1.68
3/3/2001	NSW	Autumn	23.0666664	0.0154	0.503350336	126.462815	9.344296126	3.580806255	1.92
3/4/2001	NSW	Autumn	22.5970704	0.0629	0.503551064	123.618531	9.142151851	5.130313151	8.36
3/5/2001	NSW	Autumn	22.2639871	0.0938	0.521986927	122.590034	9.046004902	6.104190629	6.42
3/6/2001	NSW	Autumn	22.4570422	0.154176	0.563481684	123.516032	8.790426445	5.566320245	16.34
3/7/2001	NSW	Autumn	23.1381605	0.202306	0.560820609	123.477349	8.674506335	5.394594011	4.68
3/8/2001	NSW	Autumn	23.7792837	0.33642	0.562015615	124.949674	8.283104302	6.545845019	20.84
3/9/2001	NSW	Autumn	23.9007759	0.192651	0.581465425	129.436044	7.81726084	5.68789694	4.18
3/10/2001	NSW	Autumn	24.4514461	0.122327	0.619175807	131.987019	7.513944327	2.543288998	0
3/11/2001	NSW	Autumn	24.0347302	0.201372	0.635195148	141.56751	7.503210637	1.208794973	0
3/12/2001	NSW	Autumn	20.2732422	0.186545	0.572750452	160.369259	7.412102624	3.27440231	2
3/13/2001	NSW	Autumn	19.778475	0.0148	0.474290332	151.356269	9.254013131	2.518657371	0
3/14/2001	NSW	Autumn	21.4970974	0.015	0.484759186	145.147364	8.85545092	2.487315155	37.44
3/15/2001	NSW	Autumn	23.1032272	0.0135	0.515528793	140.779796	8.552042441	3.581368409	7.98
3/16/2001	NSW	Autumn	21.5678339	0.449342	0.690499064	182.656918	7.06461404	4.077979531	2.925409228
3/17/2001	NSW	Autumn	17.9032116	0.251586	0.641763208	201.126684	6.63651944	2.989840838	0
3/18/2001	NSW	Autumn	18.2741303	0.056	0.647512542	182.144704	8.345103107	3.285506556	1.21
3/19/2001	NSW	Autumn	19.6475978	0.0173	0.651146256	168.385667	8.219579722	1.024999716	7.8
3/20/2001	NSW	Autumn	22.1376803	0.0408	0.579814199	157.287825	8.336651207	3.083045807	0
3/21/2001	NSW	Autumn	18.19684	0.446736	0.662458666	166.937224	6.27872664	3.129228312	0
3/22/2001	NSW	Autumn	18.2480132	0.0317	0.626268593	162.94852	7.70821849	3.032039203	0
3/23/2001	NSW	Autumn	20.9809391	0.0177	0.592813003	155.909001	7.994522212	1.971663026	1
3/24/2001	NSW	Autumn	21.3114341	0.0933	0.715360013	163.107496	4.849813693	2.482294082	0
3/25/2001	NSW	Autumn	20.2869473	0.297117	0.674327754	180.955783	6.304720125	2.900365109	1.32
3/26/2001	NSW	Autumn	18.8984211	0.0247	0.581330453	168.860176	8.236528617	2.841659086	0
3/27/2001	NSW	Autumn	17.9299386	0.0473	0.562666386	158.566974	8.556354588	4.182017937	3.35
3/28/2001	NSW	Autumn	16.5100695	0.0175	0.54063254	150.520132	8.380364163	5.066316388	42.4705477
3/29/2001	NSW	Autumn	17.5634371	0.0201	0.559559189	144.146139	7.962810316	3.476067622	0
3/30/2001	NSW	Autumn	19.5319901	0.018	0.544454314	140.172362	7.837811809	2.647859288	7.25
3/31/2001	NSW	Autumn	20.6557291	0.01	0.53439558	135.947121	7.772911378	2.399603395	14.74
4/1/2001	NSW	Autumn	20.3475783	0.0183	0.545675955	132.284221	7.585496036	0.598376142	5.06
4/2/2001	NSW	Autumn	19.2793526	0.0468	0.541121764	131.250797	7.335603974	3.813080756	6.48
4/3/2001	NSW	Autumn	18.736903	0.0275	0.549500583	131.328659	7.082598101	1.727558868	10.44
4/4/2001	NSW	Autumn	19.2509452	0.00288	0.50906609	128.671835	7.811589699	0.780215585	10.28

Region

NSW
NT
QL
SA
TA
VI
WA

Season

Winter
Spring
Summer
Autumn

Crowdcast April 5, 2021 – Predict Wildfires in Australia

- Welcome and Call for Code Overview (Upkar)
- Overview & Summary Results (Susan)
- The datasets (Hendrik)
- A technical overview of the models from 3 teams
  - yau_yee_Italy Team (Marco)
  - Data Warriors (Albert)
  - NA Team (Brianna, Ned)
- A summary comparison of the modeling approaches (Wiktor)
- Students using the challenge as part of their university program
  - Dr Manar, Emad, Abdelhady
- What's next (Susan)
- Call for Code Close and Q&A (Upkar)

What's Next

Session on April 26 - Prometeo – Firefighters Health Platform, a Call for Code Open Source Project

- Register : https://www.crowdcast.io/e/prometeo-a-call-for-code/
- A prototype sensor which sends environmental telemetry processed by AI to monitor fire fighter health risk.

Call for Code Global Challenge 2021

- Register here: https://callforcode.org/global-challenge/
- If you are a student remember to say that you are when you register
- If you are an IBMer go to the IBM internal slack channel #call-for-code

Pull together the resources created for the Wildfires challenge – & publish as resources for the global call for code

- https://github.com/Call-for-Code/Spot-Challenge-Wildfires

Call for Code Global Challenge 2021

Calling all students: Call for Code Global Challenge 2021. An excellent activity for students seeking tech for good projects (& prizes). This year's themes include clean water and sanitation, zero hunger, responsible production, green consumption. Please share with student clubs, enthusiasts, colleagues

- Register here: https://callforcode.org/global-challenge/ If you are a student please indicate in registration. You will then start receiving invitations to activities to help you along
- Watch the kick-off event: https://www.youtube.com/watch?v=Pcw24VWeCRM *(Mami Mizutori United Nations Office for Disaster Risk Reduction, Chelsea Clinton & more)*
- Get started

Materials created as we went along

Useful Links

- The Contest landing page http://ibm.biz/cfcsc-wildfires
- The Contest GitHub https://github.com/Call-for-Code/Spot-Challenge-Wildfires
- The Contest leaderboard http://ibm.biz/cfcsc-wildfires-lead
- Education materials to get started with the Call for Code Spot Challenge for Wildfires
- The Finale : Call for Code Spot Challenge for Wildfires
- Call for Code Spot Challenge for Wildfires Predictions: Comparing approaches
- Slack Workspace http://callforcode.org/slack Channel #cfcsc-wildfires
- Helpful blog https://medium.com/ibm-data-ai/predicting-australian-wildfires-with-weather-forecast-data-8d1cc983c863

Office Hours - From Dec 15 2020 – Jan 26, 2021

Crowdcasts

- Nov 23 about notebooks and exploring the datasets https://www.crowdcast.io/e/call-for-code-spot/
- Nov 30 about datasets and how to submit to leaderboard https://www.crowdcast.io/e/call-for-code-spot-2/
- Dec 7 about how to use AutoAI with the data https://www.crowdcast.io/e/call-for-code-spot-3
- Jan 11 introducing the Contest and Leaderboard https://www.crowdcast.io/e/call-for-code-spot-4/
- Jan 18 hear from Team WildFireNet who participated in another hackathon https://www.crowdcast.io/e/call-for-code-spot-5
- Jan 25 to hear about Hypothesis Testing for Time Series https://www.crowdcast.io/e/call-for-code-spot-6
- Feb 22 to hear from IBMers building models for the wildfires challenge https://www.crowdcast.io/e/call-for-code-spot-7
- Apr 5 to hear from the leading teams https://www.crowdcast.io/e/predicting-australian/ and from students at the University of Sharjah using the spot challenge as a case study
- Apr 26 to hear from Prometeo – Firefighters Health Platform, a Call for Code Open Source Project https://www.crowdcast.io/e/prometeo-a-call-for-code/
- Join Call for Code Global Challenge 2021 - Register : https://callforcode.org/global-challenge/
    - *This year's themes include clean water and sanitation, zero hunger, responsible production, green consumption*