Regional Workshop on
The Development of Transport Statistics and Transport
Related Sustainable Development Indicators in the Arab
Countries
Beirut, Lebanon: 16-18 September 2019

ESCWA Pilot Study On
Alternative And Complementary
Data Sources On Road Safety:
Lessons Learned

Mohamad Hossary
Economic Statistics Section (ESS),
SD- ESCWA
Agenda

• Introduction
• Typical Data Sources
• The Need to Link Different Sources
• ESCWA’s Pilot Project on Selected Countries
• The Types of Data
• The Challenges
• The Analysis
• The Way Forward
The Status of Road Safety

• Risk of road crashes in low to middle income countries is more than 3 times higher than high income countries (1)

• Average rate of 27.5 deaths per 100,000 population vs 8.3 deaths (1)

• Traffic related injuries are leading cause of death for children and young adults (1)
The Typical Data Source

- Official Data usually is taken from police records

<table>
<thead>
<tr>
<th>Attribute name</th>
<th>Attribute description</th>
</tr>
</thead>
<tbody>
<tr>
<td>psn_id</td>
<td></td>
</tr>
<tr>
<td>record_status</td>
<td></td>
</tr>
<tr>
<td>acd_date</td>
<td>Accident Date</td>
</tr>
<tr>
<td>acd_time</td>
<td>Accident time</td>
</tr>
<tr>
<td>acc_location</td>
<td>statement describes the location of the accident</td>
</tr>
<tr>
<td>id</td>
<td>System Generated ID for the Accident</td>
</tr>
<tr>
<td>acc_type</td>
<td>Accident category (i.e. minor, major, etc.)</td>
</tr>
<tr>
<td>acc_cause</td>
<td>The cause of the accident</td>
</tr>
<tr>
<td>weather</td>
<td>Description of the weather</td>
</tr>
<tr>
<td>road_status</td>
<td>Description of the road status</td>
</tr>
<tr>
<td>age</td>
<td>Main actor age</td>
</tr>
<tr>
<td>gender</td>
<td>Main actor gender</td>
</tr>
<tr>
<td>injury_severities</td>
<td>Driver Injury level</td>
</tr>
<tr>
<td>driving_license_issue_date</td>
<td>DRIVING_LISCENCE_ISSUE_DATE</td>
</tr>
<tr>
<td>occupation</td>
<td>Driver occupation</td>
</tr>
<tr>
<td>intoxication</td>
<td>INTOXICATION</td>
</tr>
<tr>
<td>seat_belt_status</td>
<td>Flag</td>
</tr>
<tr>
<td>year_manufactured</td>
<td>YEAR_MANUFACTURED</td>
</tr>
<tr>
<td>insurance_company_name</td>
<td>INSURANCE_COMPANY_NAME</td>
</tr>
</tbody>
</table>

The Issues With Police Data

• Not all crashes are reported to police \(^{(2)}\)
• Not all countries have 30-day follow up rule for injuries \(^{(2)}\) \(^{(3)}\)
• Main purpose of data is for legal proceedings
The Need to Link Different Sources

Help overcome some of the issues (health and police data)

Additional features
ESCWA’s Pilot Project on Selected Countries

- Selected sector: transport
- Lebanon is one of the countries selected
- Objective is to gain insights into the factors correlated to fatal car crashes
The Different Data

- Official Police Data
- Weather and Solar Azimuth
- Road Related Data
- Traffic Related Data
- Health Related Data
Official Police Data

- Available from police and ministries of interior
- Composed of reports from crash sites
- Was requested and still pending from countries
Weather and Solar Azimuth

- Position of the sun in the sky
- Visibility
- Precipitation
- Temperature
- Available online through scraping or api
- May cost money if api is used
Road Related Data

- Road Maintenance Records
- Satellite Imagery
- CCTV footage
- Maps
Road Related Data

Road Maintenance Records
- When each road was serviced
- Predict the condition of the road

Satellite Imagery
- Road Quality (if highway, and image is high resolution)
- Shape of the road

CCTV footage
- Detect issues with the road
- Shape of the road
- Obstructions

Maps
- Road shape files
- Extract road shapes and curvatures
Road Related Data

- **Road Maintenance Records**
  - Available at the ministry of public works

- **Satellite Imagery**
  - Can be found free from NASA and ESA
  - Can be found paid from other providers for tens of thousands of dollars

- **CCTV footage**
  - Available from police
  - Available from ministry of transportation

- **Maps**
  - Available free from Openstreet maps
  - Available paid from Here Maps, Google Maps, and other commercial mapping companies
Traffic Related Data

CCTV
- Number of Vehicles
- Average Speed

Navigation
- Average Speed
- Historical Speed

Government Counts
- Number of vehicles counted then extrapolated for the year

Sensor Data
- Vehicle Counts
- Speed on road segment
Traffic Related Data

CCTV
- Available from police
- Available from ministry of transportation

Navigation
- Available from Google Maps or Here Maps
- Expensive can range from 10,000$ to more than 100,000$

Government Counts
- Available from NSO or ministry of public works

Sensor Data
- Available from ministry of transportation or public works
The Challenges: Obtaining the Data

• In the Middle East there is a general unwillingness to share data
• Issue of availability, reliability, and Continuity of data in open data hubs
• Data Silos and NSOs being unaware of what data the country had
The Challenges: Data and Analysis

• Mobile phone data was discarded due to challenges in linking to road segments
• Large files and certain algorithms require storage and hefty compute performance
The Analysis

1. Define the target
2. Preprocess the alternate data
3. Consolidate the data into one dataset
4. Clean and preprocess the main dataset
5. Conduct descriptive statistics
6. Run ML algorithms and ensembles
7. Extract feature importance using different python packages
Processing the Alternate Data

- CCTV Data
- Satellite Imagery
- Weather and Solar Azimuth
Processing the Alternate Data

CCTV Data

Satellite Imagery

YOLO v3 network Architecture

Ayoosh Kathuria, What's new in YOLO v3?, from https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b, 9/14/2019
Processing the Alternate Data

CCTV Data

Satellite Imagery

Ayoosh Kathuria, What's new in YOLO v3?, from https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b, 9/14/2019
Processing the Alternate Data

**CCTV Data**
- Convolutional Neural Network (CNN) to extract average traffic speed, traffic counts, and road quality

**Satellite Imagery**
- CNN to extract road shape and road quality (if road was wide enough)

**Weather and Solar Azimuth**
- Web scrape or access through api to extract visibility, location of the sun, precipitation, and other weather related data
Analysis Through Machine Learning

- Consolidate data
- Merge alternate data and official police data

- Clean and feature engineer data
- Clean the data and extract new features from existing ones

- Implement machine learning models
- Implement several models and evaluate through F1 score.
  Join models in ensemble to improve score if needed
Making ML Interpretable

• Permutation importance
  – Randomly change the value of a column keeping all else the same
  – Feature is important if accuracy drops after change

• Partial Dependence Plots
  – Shows the effect of a feature on prediction
  – Similar technique to permutation importance
Conclusion and the way forward

- Official data has some issues that need to be addressed
- Alternative data can be expensive
- Limited access to alternative data
- Different data and tools will be used going forward to address these issues
References


3. Traffic Safety Statistics between Project Partner Countries (Egypt, Jordan and Lebanon, Poland, Spain and Sweden) and Internationally (2015), MENASAFE