אלגוריתמים מבוססים למידת מכונה להעשרת OpenStreetMap בנתונים חיוניים לחישוב מסלולים להולכי רגל עיוורים

אחיטוב כהן

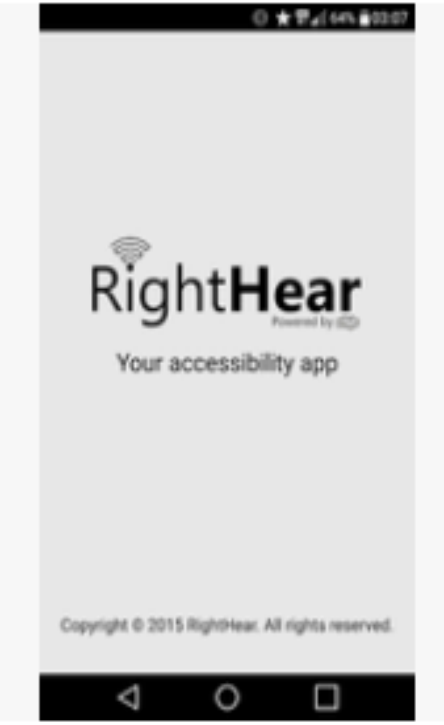
מנחה: ד"ר שגיא דליות

Technion
Israel Institute of Technology

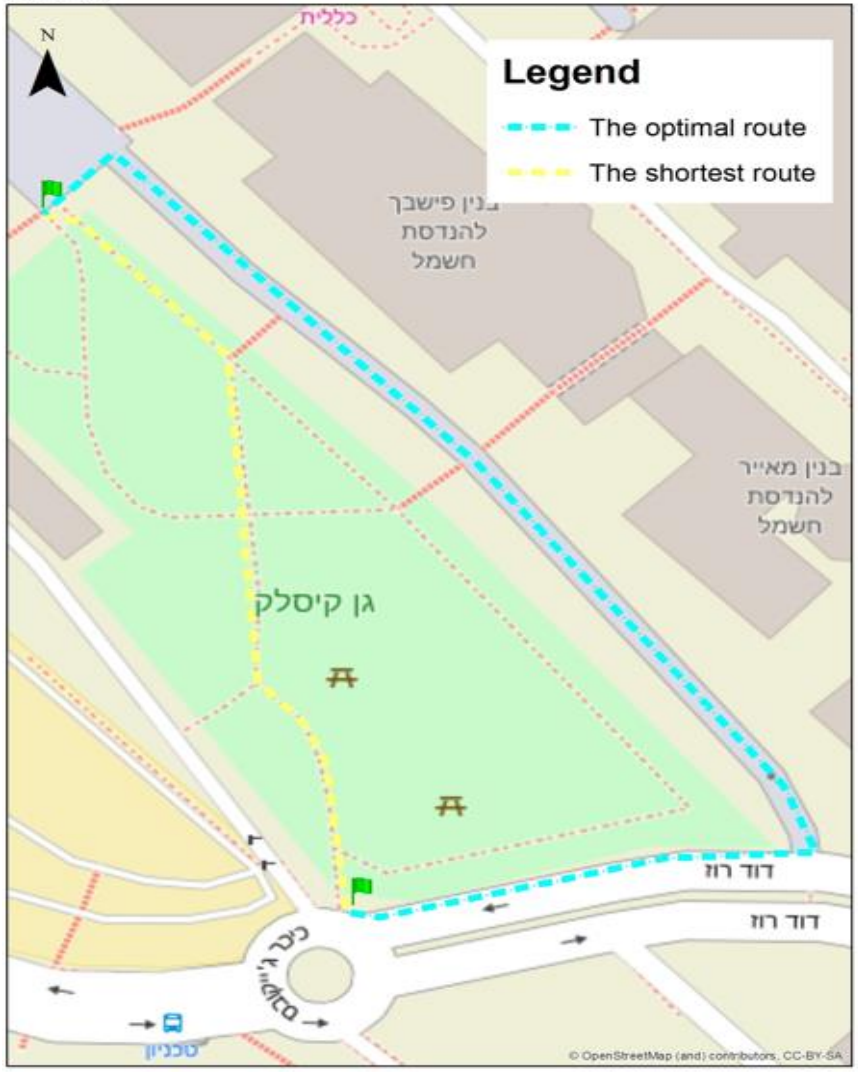
ECSL
Environmental Crowdsourcing Lab

MAPPING & GEO INFORMATION ENGINEERING

נכון לשנת 2018, חיים בעולם כ- **441.5 מיליון** אנשים **לקויי ראייה**;
36 מיליון מתוכם **עיוורים לחלוטין**, והשאר עם לקות ראייה בינונית עד גבוהה (ארגון הבריאות העולמי).



Haifa


Legend
The optimal route
The shortest route
N
בנין פישבך להנדסת חשמל
בנין מאייר להנדסת חשמל
גן קיסלק
דוד רוז
דוד רוז
כיכר הקפה
טכניון
© OpenStreetMap (and) contributors, CC-BY-SA

NYC

Legend
The optimal route
The shortest route
N
Wagamama
5th Avenue
Admiral David Farragut
Chester Alan Arthur
Moira Ann Smith Playground
Madison Square Park
Mad. Dogs
Broadway
Madison Avenue
Shake Shack
Roscoe Conkling
East 23rd Street
Eleven Madison Park
Met Life Tower
Metropolitan Life
Charles
© OpenStreetMap (and) contributors, CC-BY-SA

הגדרת הבעיה

Aluminum St
Google

OSM

Aluminum St
Aluminum St
Aluminum St
Smadar St
Ha-Shalva St
Google Maps

Gov Map

הגדרת הבעיה

מידע עירוני

- רעש
- רמת צפיפות הולכי הרגל ברחובות
- חניה של רכבים על מדרכות
- עבודות תחזוקה על מדרכות
- חסימות
- ריח

מידע מרחבי

- מדרכות
- שבילי הליכה
- מעברי חצייה
- ציוני דרך, כגון עצים, גדרות, פחים

מטרות המחקר

זיהוי מאפיינים במרחב, שלהם קורלציה לקיומן של התופעות הנדרש
והבנת הקשר בי

חיזוי, בצורה מדויקת, של התופעות ועוצמתן ליצירת נתונים חדשים, וזאת בעזרת מודלים מבוס
אלגוריתמי למידת מ

בחינת השפעת נתוני המיפוי החדשים על איכות התוצאות המתקבלות מאלגוריתמי חישוב המסלול להולכי רגל עיוורים.

חיזוי וחישוב צפיפות הולכי הרגל

Transportations

tinycards.duolingo.com

Tourism

www.thisisinsider.com

Offices

electra.co.il

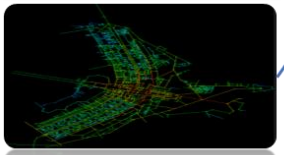
Aesthetics

nparks.gov.sg

Shops

iflychs.com

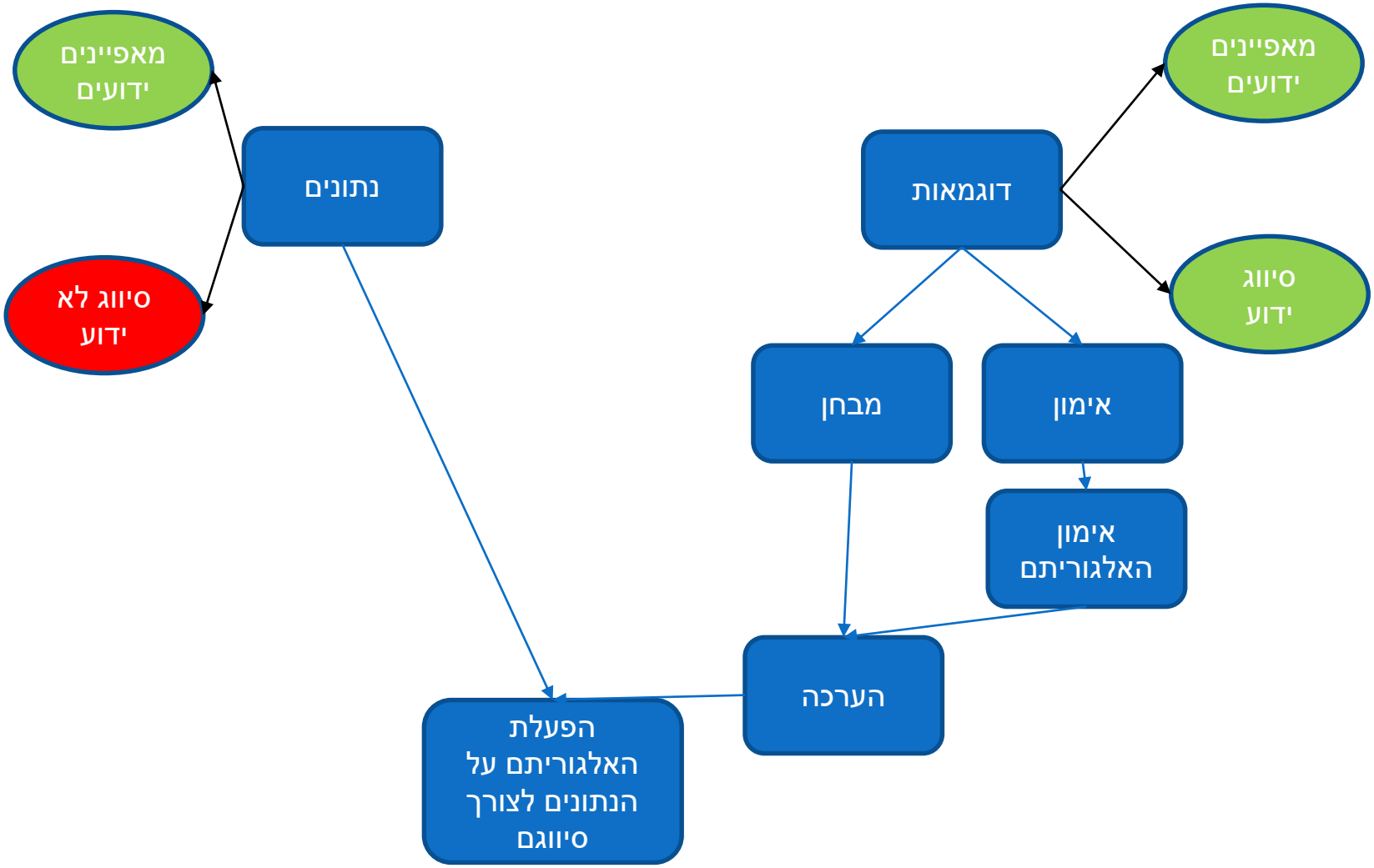
Space syntax



wikipedia.org

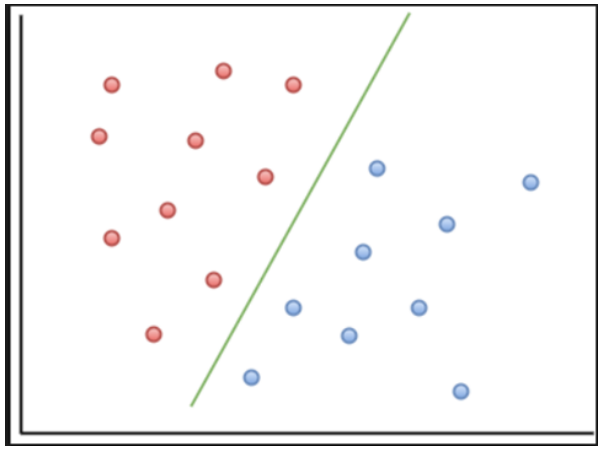
?

השיטה לחיזוי וחישוב - למידת מכונה

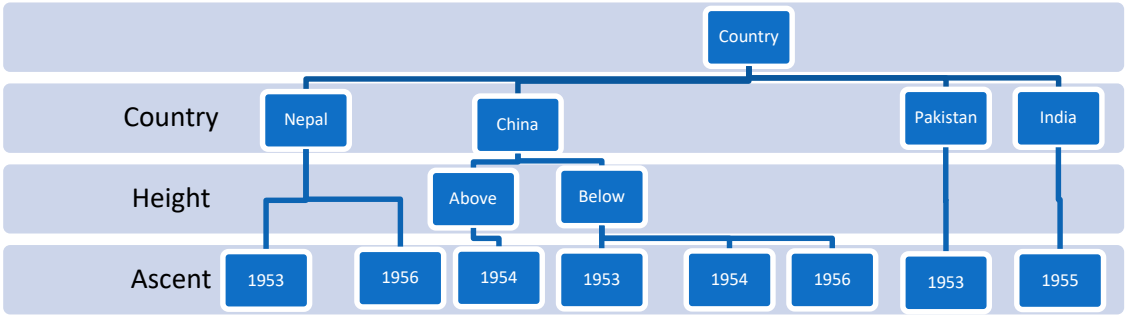


השיטה לחיזוי וחישוב - למידת מכונה

Support Vector Machine (SVM)



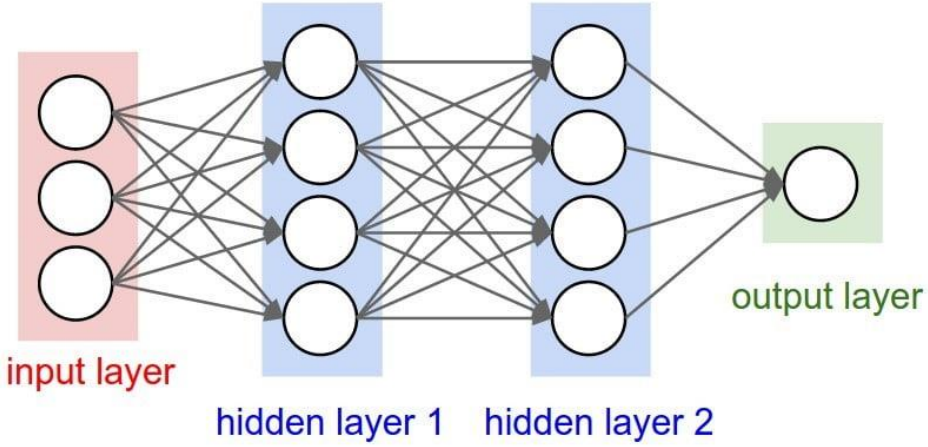
Decision tree



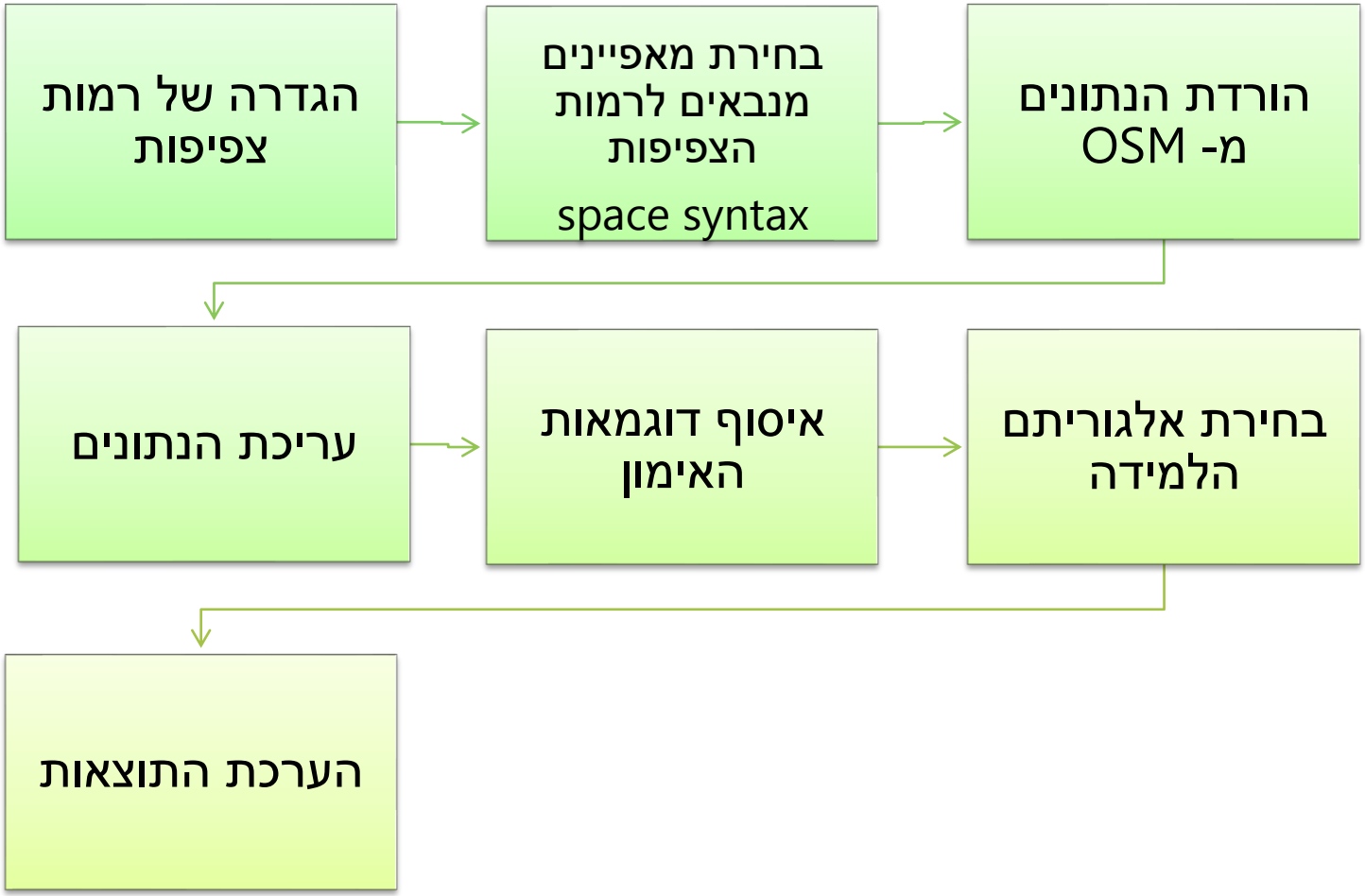
Naive Bayes classifier

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

ANN

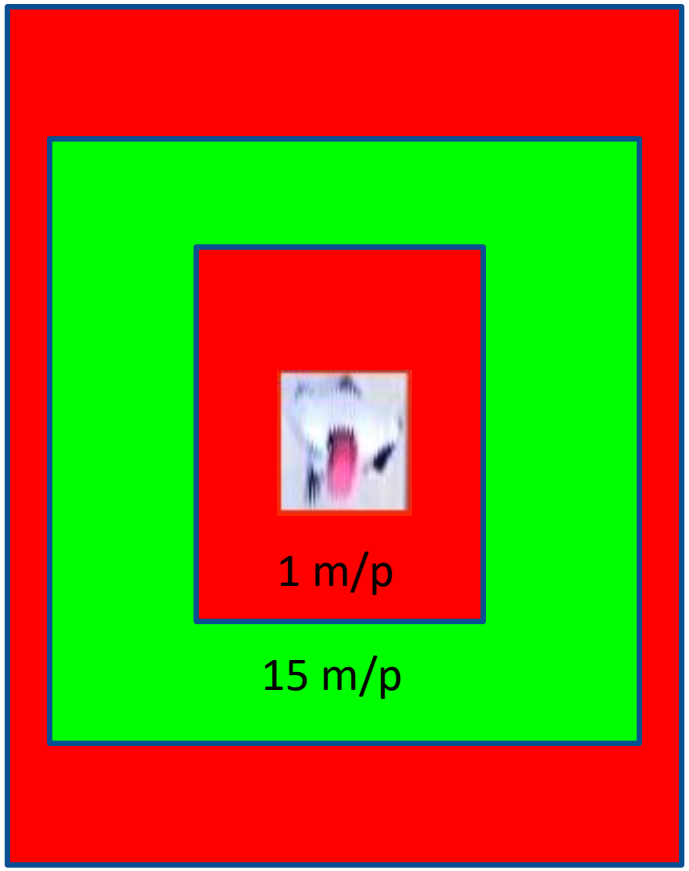


חיזוי וחישוב צפיפות הולכי הרגל

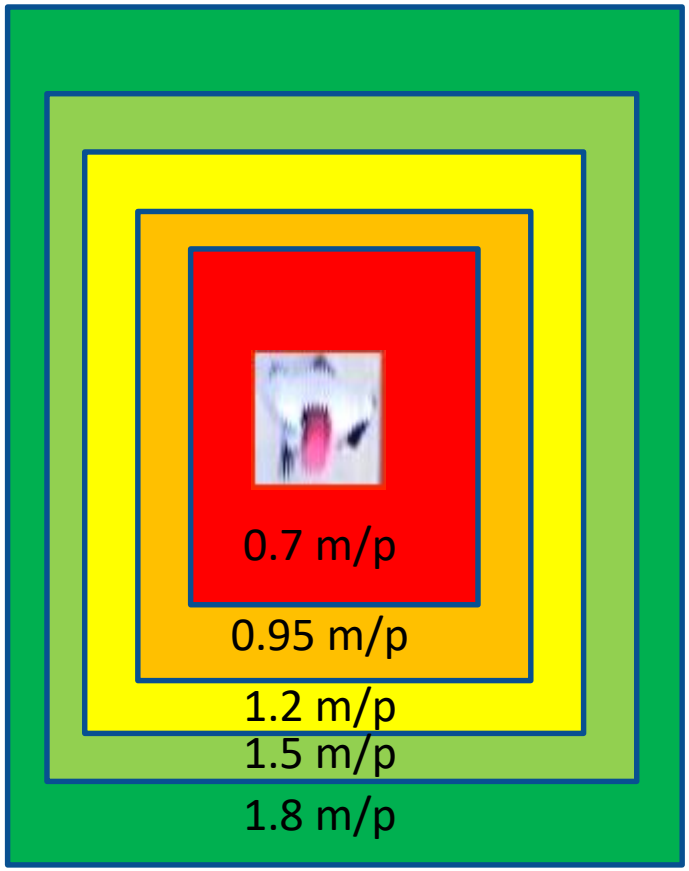


הגדרה של רמות צפיפות

רמות צפיפות עבור הולכי רגל עיוורים



רמות צפיפות עבור הולכי רגל



"Pedestrian, Crowd and Evacuation Dynamics", Helbing and Johansson ,2009

בחירת מאפיינים מנבאים לרמות הצפיפות

Space Syntax
- integration
- choice

Tourism
- points
- polygons

Shop
- points
- polygons

Amenity
- points
- polygons

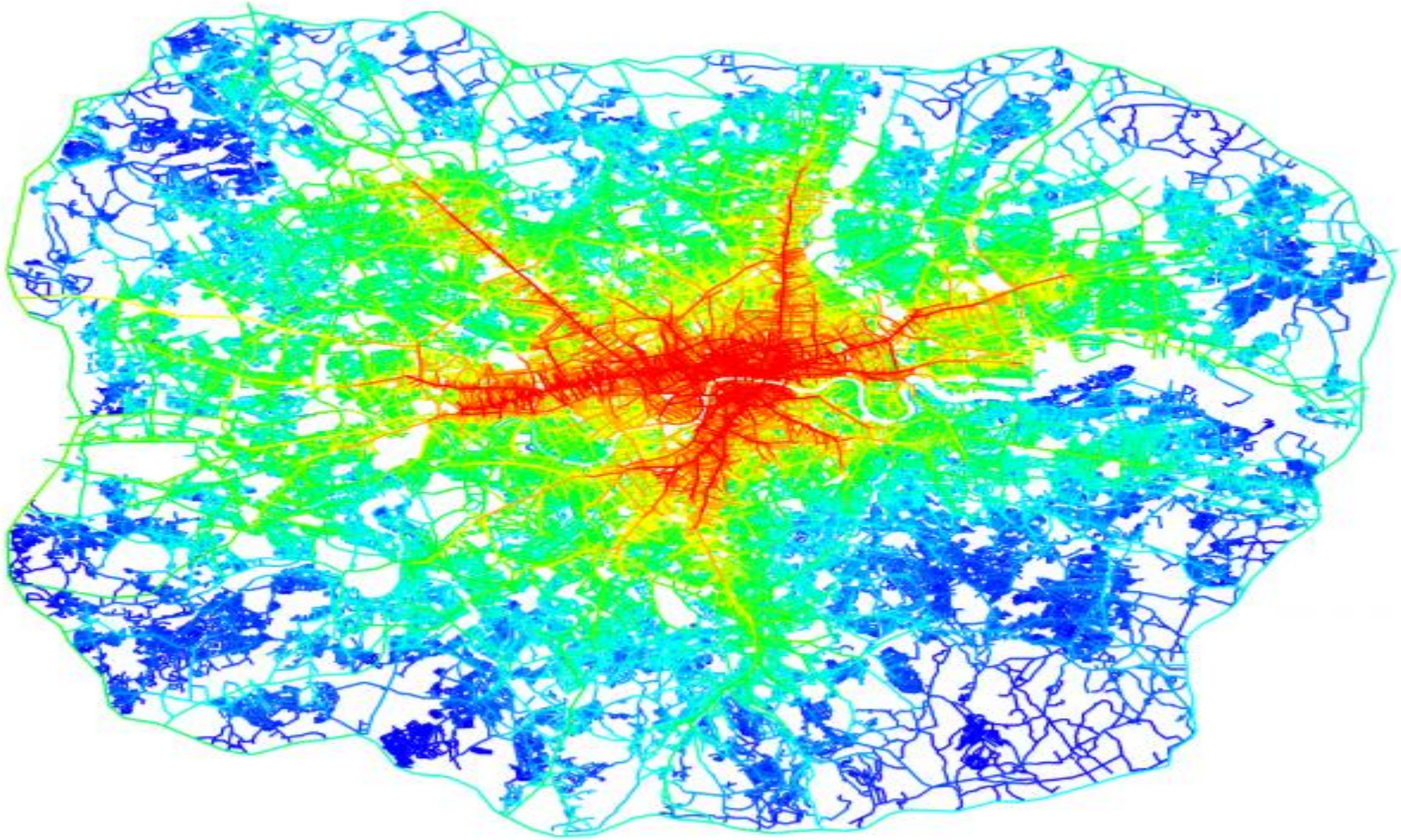
Office
- points
- polygons

Aesthetics
- points
- polygons

Transportation
- points
- polygons

Time

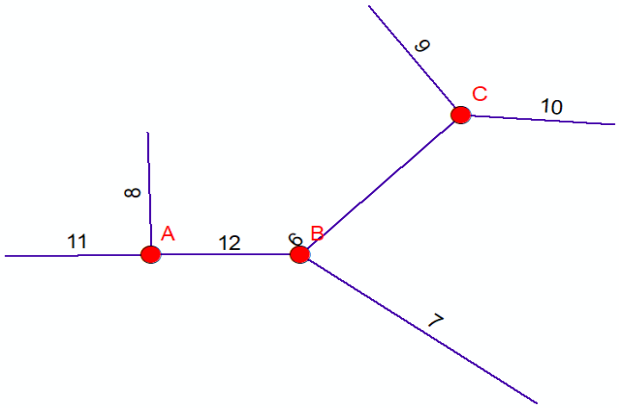
Space Syntax



http://otp.spacesyntax.net/overview-2/

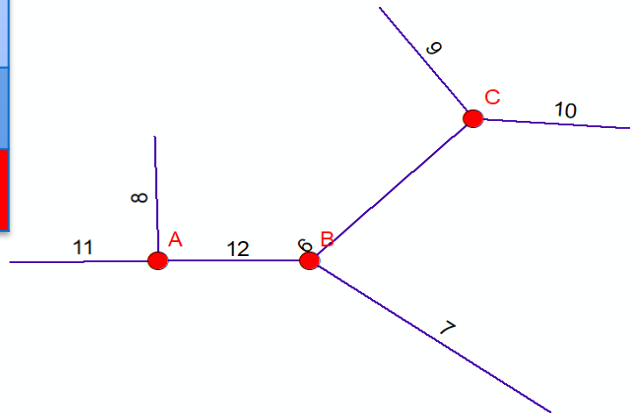
Integration - Space Syntax

	11	8	12	6	7	9	10	INT
11	0	1	1	2	2	3	3	12
8	1	0	1	2	2	3	3	12
12	1	1	0	1	1	2	2	8
6	2	2	1	0	1	1	1	8
7	2	2	1	1	0	2	2	10
9	3	3	2	1	2	0	1	12
10	3	3	2	1	2	0	0	12

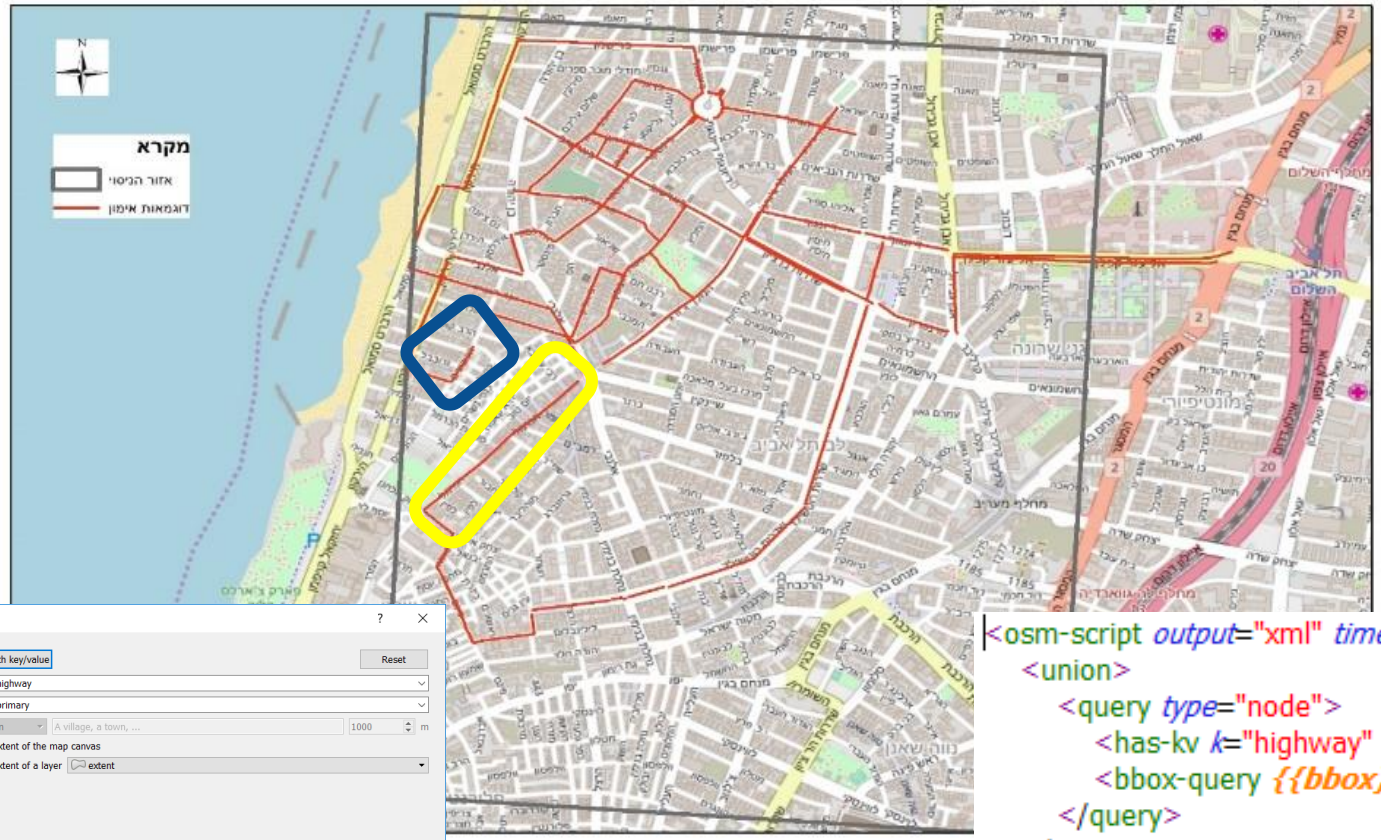


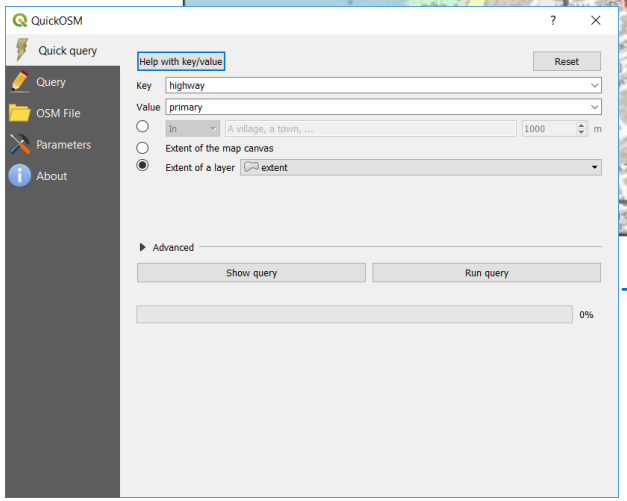
Choice - Space Syntax

	11	8	12	6	7	9	10
11	1	0.5	0.5	0.25	0.25	0.125	0.125
8	0.5	1	0.5	0.25	0.25	0.125	0.125
12	0.25	0.25	1	0.25	0.25	0.125	0.125
6	0.125	0.125	0.25	1	0.25	0.25	0.25
7	0.25	0.25	0.5	0.5	1	0.25	0.25
9	0.125	0.125	0.25	0.5	0.25	1	0.5
10	0.125	0.125	0.25	0.5	0.25	0.5	1
CH	2.375	2.375	3.25	3.25	2.5	2.375	2.375



הורדת הנתונים מ- OSM

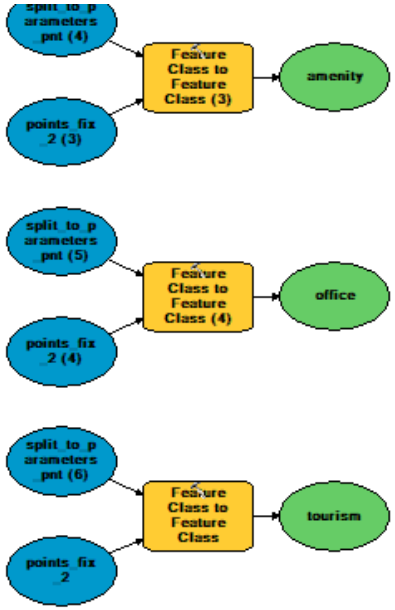


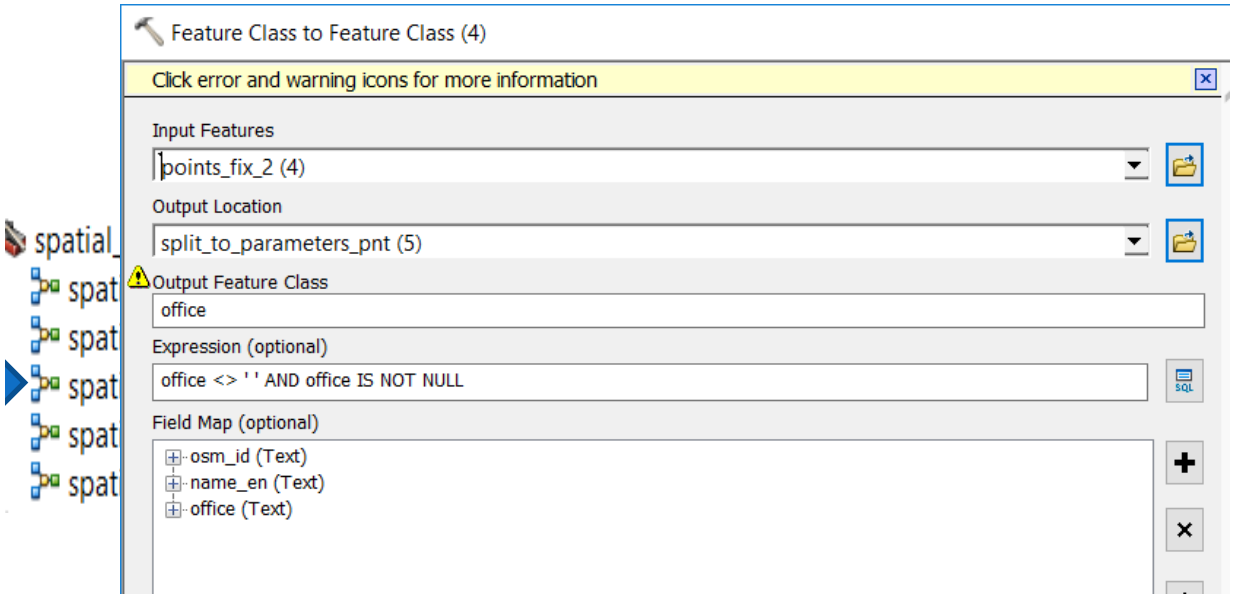



```
<osm-script output="xml" timeout="25">
  <union>
    <query type="node">
      <has-kv k="highway" v="bus_stop"/>
      <bbox-query {{bbox}}/>
    </query>
  </union>
  <union>
    <item/>
    <recurse type="down"/>
  </union>
  <print mode="body"/>
</osm-script>
```

הורדת הנתונים מ- OSM

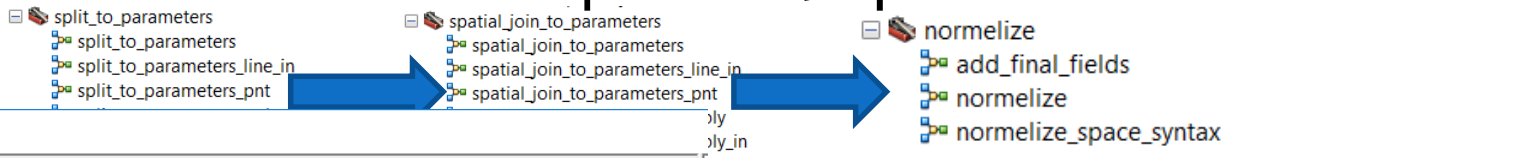
עריכת הנתונים

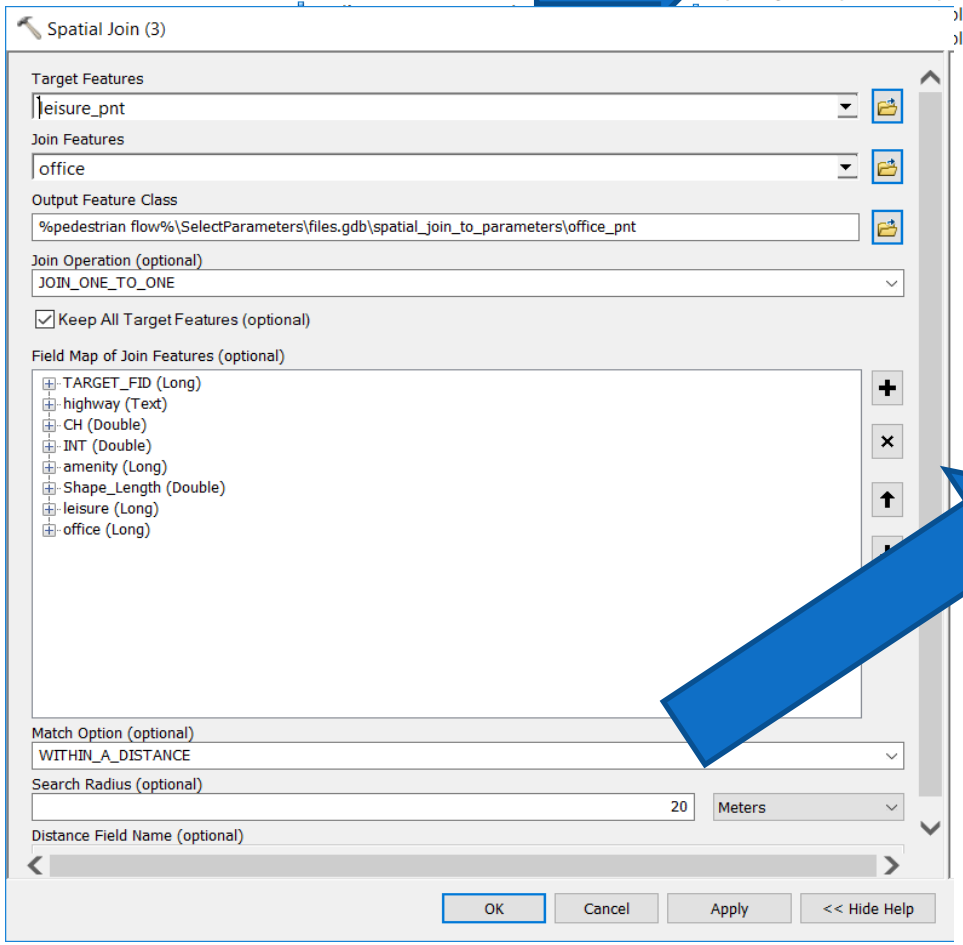




עריכת הנתונים

הוספת המאפיינים לכל דרך על בסיס קירבה





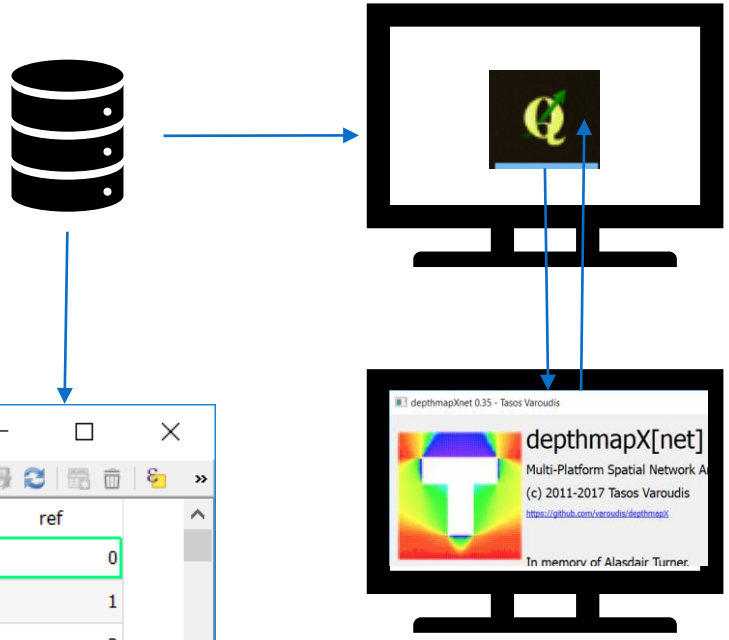
SHARE_A_LINE_SEGMENT_WITH

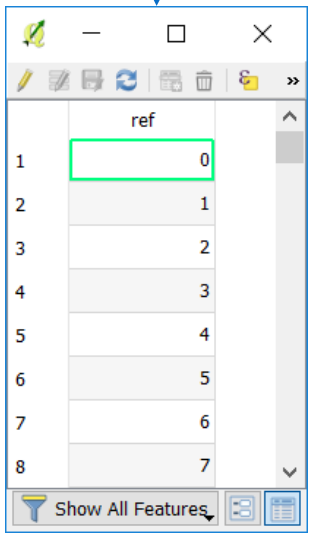
WITHIN_A_DISTANCE – 20 meters

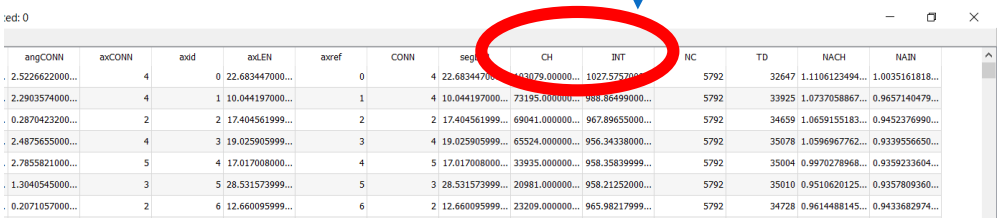
WITHIN_A_DISTANCE – 20 meters

INTERSECT

תוצאות וניתוח ראשוני – עריכת המידע







איסוף דוגמאות אימון

- בוצע בשלושה ימים בין השעות 8:30 בבוקר ל- 19:00 בערב.
- נבחרו כ- 60 קטעי רחוב בתל-אביב, שנמדדו בין פעם אחת לשלוש פעמים ביום בשעות שונות.
- חלק מהרחובות שנבחרו היו רחובות ראשיים במרכזי תעסוקה ואזורי מסחר ובילוי, וחלק רחובות צדדיים ושקטים.
- 80% מהדוגמאות (מסקר השדה) היוו קלט לאימון ומידול ה- ML (נתוני אימון), ו- 20% הנותרים נועדו לבחון את האלגוריתם (נתוני בחינה).
- איסוף ושמירת דוגמאות האימון התבצעה בעזרת אפליקציית collector של ArcGIS.

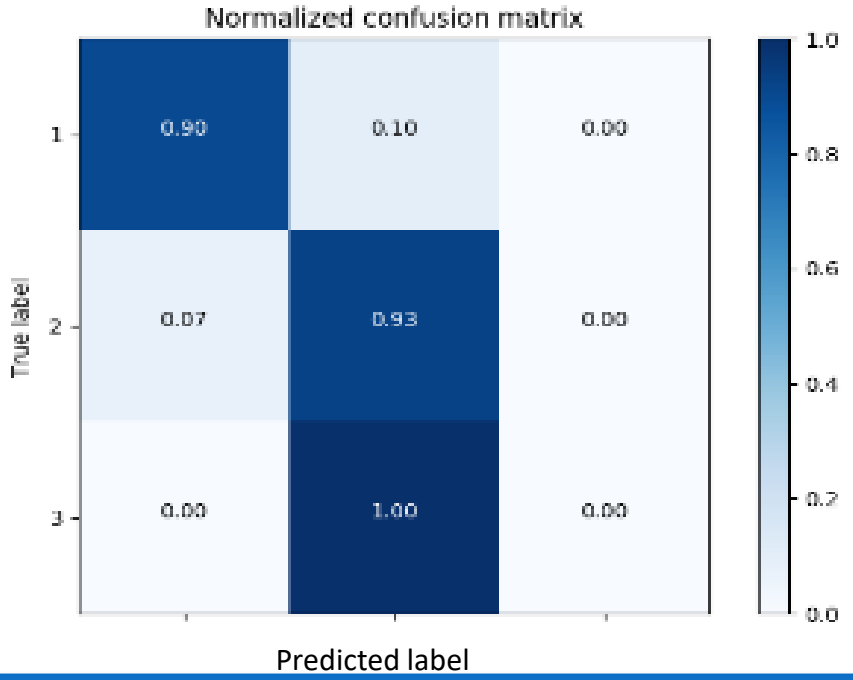
איסוף דוגמאות אימון

סה"כ	דוגמאות מבחן	דוגמאות אימון	רמת צפיפות (label)
52	10	42	**1**
71	14	57	**2**
2	1	1	**3**

בחירת אלגוריתם הלמידה

אלגוריתם	דיוק הניבוי ביחס לסט המבחן (ערכים בין 0 ל- 1)
MultinomialNB	0.44
AdaBoost	0.60
Gaussian Naive Bayes	0.60
BernoulliNB	0.64
GradientBoosting	0.64
MLPClassifier	0.64
DecisionTreeClass	0.80
Artificial Neural Network	0.88

תוצאות



Traffic Flow	Precision	Recall	F1-score	Support
Low	0.90	0.90	0.90	10
Fine	0.87	0.93	0.90	14
High	0.00	0.00	0.00	1

הערכת התוצאות

הרצת האלגוריתם על כל הרחובות במרכז תל אביב בשעה 12:00 בצהריים לקבלת מדד צפיפות להולכי רגל עיוורים.

הערכת התוצאות

ההשפעה של המאפיינים על הדיוק

המשקלים של כל מאפיין על השכבה הראשונה

ללא המאפיין	דיוק הניבוי ביחס לסט המבחן - points	דיוק הניבוי ס - p
Space Syntax	Integration – 0.8	C
Tourism	0.72	
Shop	0.76	
Amenity	0.72	
Office	0.6	
Aesthetics	0.6	
Transportation	0.8	
Time	0.68	

Time	0.96
Shop polygon	0.36
Aesthetics polygon	0.17
Amenity polygon	0.36
Transportation point	0.29
Tourism point	0.45
Shop point	0.28
Office point	0.38
Aesthetics point	0.02
Amenity point	0.47
CH	1.60
INT	1.04

	דיוק הניבוי ביחס לסט המבחן - points
Syntax	0.72
סביבתיים	0.8

סיכום ומסקנות

- למרות מספר דוגמאות האימון הקטן, המודל שנבנה בעזרת ANN מצליח לחזות את רמות הצפיפות השונות ב- 90% מהמקרים.
- נראה כי המודל מסתמך בעיקר על נתונים גאומטריים שקיימים ב- OSM – רשת הכבישים (בערים) ששלמות המיפוי שלהן ב- OSM טובה ועדכנית, בניגוד לישויות נקודתיות, כגון חנויות, ששלמות המיפוי שלהן לא תמיד כך.
- בכדי לקבל תוצאות מהימנות יותר:
  - התבססות על רשת הדרכים המבוססת על שבילים/מדרכות להולכי רגל בלבד.
  - איסוף דוגמאות אימון נוסף רחב ממספר ערים, ולא רק ממרכז תל אביב.
  - את ה- ANN מומלץ להפעיל על מספר גדול יותר של דוגמאות אימון.
  - כמו שהוצג במתודולוגיה, יש להפעיל מידול מתוך אלגוריתמים חלופיים, ולמצוא את זה האופטימלי.
  - השוואת המודל שהתקבל, המתבסס על נתוני OSM, למודל שיתקבל המתבסס על מאגר נתונים אחר.

Thank you