

## PROGETTI CONCLUSI 2019 – Il Call

<b>Titolo progetto</b>	<b>“Classificazione avanzata del verde urbano da dati Open e Telerilevati”</b>
<b>Descrizione</b>	Il progetto ha lo scopo di mappare e georiferire puntualmente sia su scala comunale sia regionale le singole porzioni di territorio adibite a verde associando a queste ultime la classe di appartenenza. Le informazioni sono state estratte da dati provenienti da diversa fonte: dati Istat per le sezioni di censimento e località abitate, da Carte Tecniche Regionali (CTR), per cogliere alcuni aspetti principali del territorio come aree verdi e strade, dal Ministero dell’Ambiente per le Aree naturali protette, dalla piattaforma OpenStreetMap per le aree dei parchi urbani, delle aree ricreative, delle aree boschive, ecc. e da immagini satellitari e ortofoto per la presenza di vegetazione sul territorio. Per la classificazione automatizzata delle aree è stato utilizzato il metodo di machine Learning (deep learning).
<b>Obiettivi</b>	L’obiettivo del progetto è potenziare l’informazione disponibile sulle aree verdi, tutti gli strumenti utilizzati e le informazioni raccolte, infatti, servono ad accrescere la qualità del dato statistico sia in termini di processo, sia di prodotto, fornendo un’informazione statistica efficiente ed efficace. L’estrazione di tutte le aree verdi presenti all’interno dei capoluoghi, a prescindere dalla gestione, porterà ad ampliare l’informazione disponibile prendendo in considerazione anche le aree verdi di proprietà privata. L’elaborazione di queste informazioni permetterà il calcolo degli indicatori per il monitoraggio del raggiungimento dei target di <i>Sustainable Development Goals</i> (SDGs) del Goal 11 dedicato alle città sostenibili. Tali indicatori sono sempre più richiesti a livello nazionale e Internazionale allo scopo di valutare la qualità della vita dei cittadini e la loro sostenibilità per il miglioramento delle politiche locali.
<b>Metodologia</b>	<p>L’estrazione delle informazioni sono state elaborate prendendo in considerazione due metodologie diverse a seconda della fonte dei dati utilizzata:</p> <ul style="list-style-type: none"> <li>- per la parte open data sono state estratte le mappe e i geodati per tutti i comuni capoluogo. La mappa a livello comunale è stata costruita prendendo in considerazione l’incrocio di tutte le informazioni degli shape file scaricabili dalle fonti sopra citate (come ad esempio parchi urbani, aree boschive, verde attrezzato, cimiteri, aree naturali protette) e le interconnessioni e sovrapposizioni esistenti (scala 1:5.000). A seguire sono state localizzate le varie tipologie di verde su mappa georiferita a livello comunale utilizzando sia strumenti web-based per l’estrazione dei dati da OpenstreetMap (overpass turbo) che le piattaforme geografiche (ARCGIS/QGIS) per l’analisi dei dati. La verifica puntuale delle informazioni estratte è stata realizzata con il software Google Earth che ha permesso di navigare all’interno di ogni area selezionata esaminando le aree verdi presenti. Per la validazione finale dei dati si sono confrontate le superfici totali delle aree georiferite estratte, con gli analoghi dati raccolti dalla rilevazione Istat, Dati ambientali nelle città (Codice IST-00907).</li> <li>- Per la parte dei dati Telerilevati e per la segmentazione delle immagini (localizzazione dei contorni delle aree coperte da vegetazione) è stata utilizzata una peculiare architettura di deep-learning; l’architettura presa in considerazione U-net, è tra le tecniche più innovative per la segmentazione delle immagini. Per addestrare questo algoritmo di tipo ‘supervised’, è stato creato uno specifico dataset costituito da immagini satellitari dell’agenzia spaziale europea ESA ‘Sentinel 2’ e dalla carta di uso del suolo della regione Puglia scelta per testare l’algoritmo. Delle 13 bande spettrali che caratterizzano le immagini sono state selezionate quattro aree con miglior risoluzione spaziale (verde, rosso, blu, infrarosso). Successivamente sono state estratte aree di training di 128*128 pixel con risoluzione a terra di 10 metri. Per ogni immagine satellitare la procedura costruita genera una seconda immagine (maschera) che raggruppa i pixel in cluster secondo 3 classificazioni: 1=conifere, 2=latifoglie e 0=altro;</li> </ul>

## Risultati ottenuti

specificare l'impatto sulla produzione statistica

I risultati ottenuti sono diversi a seconda della metodologia utilizzata:

- per la parte Open è stato costruito un database contenente i geodati di tutte le città capoluogo e sono state prodotte le mappe a livello comunale per più della metà dei capoluoghi (56 dei 109 capoluoghi), contenenti le informazioni delle varie tipologie di verde: grandi e piccoli parchi urbani, orti botanici, cimiteri, aree di arredo urbano, forestazione urbana, giardini scolastici, aree boschive e aree sportive. Inoltre per la confrontabilità dei dati, sono state prodotte tutte le aree verdi dei capoluoghi della regione Puglia utilizzata come test per la parte dei dati Telerilevati. Infine le informazioni sulle aree verdi estratte sono state confrontate con le stesse informazioni provenienti dal modulo del verde urbano della rilevazione Istat "Dati ambientali nelle città".
- Per la parte di dati Telerilevati è stato prodotto un dataset con oltre 11.000 tile, utile alla segmentazione di conifere e latifoglie. La procedura è stata realizzata in *Python* con l'uso di apposite librerie di processamento immagini (gdal). Dal confronto puntuale dei dati tra maschera target e maschera predetta (cella per cella), è stata ottenuta un'accuratezza della matrice degli errori (*confusion matrix*) del 97%. Questo risultato confrontato con le metriche DICE, utilizzate per la verifica delle informazioni dà una precisione media del 47%.  
Per la prima volta in Istat si sono utilizzati algoritmi di deep-learning con la creazione della rete convoluta (U-net) adatta alla segmentazione, avvalendosi di una (GPU) che permette il rendering di una grossa quantità di immagini.

Per aumentare la performance della segmentazione, una possibile soluzione è l'integrazione con altre bande satellitari che riescono a classificare anche le diverse specie di piante.

Questa procedura può essere applicata anche per la produzione di altri dataset per la segmentazione degli edifici, delle strade e di altre classi di vegetazione, ecc.