



Laboratorio Innovazione.

Risultati e prospettive

24
NOVEMBRE
2023

STIMA DEI CHILOMETRI PERCORSI DAI VEICOLI DA FONTE BIG DATA E DATABASE AMMINISTRATIVI PER LA COSTRUZIONE DI INDICATORI SULL'INCIDENTALITA' STRADALE

Marco Broccoli – DCME/MEA - Responsabile progetto

#IstatWebinar

Stima dei chilometri percorsi dai veicoli sulle strade del territorio nazionale da fonte Big data e database amministrativi per la costruzione di indicatori sull'incidentalità stradale.

- Responsabile del progetto - Marco Broccoli (DCME/MEA) - broccoli@istat.it
- Responsabile della componente IT e webscraping - Riccardo Giannini (DCIT/ITB) - rigianni@istat.it
- Responsabile per i contenuti tematici - Silvia Bruzzone (DCSW/SWC) - bruzzone@istat.it

Indice della presentazione

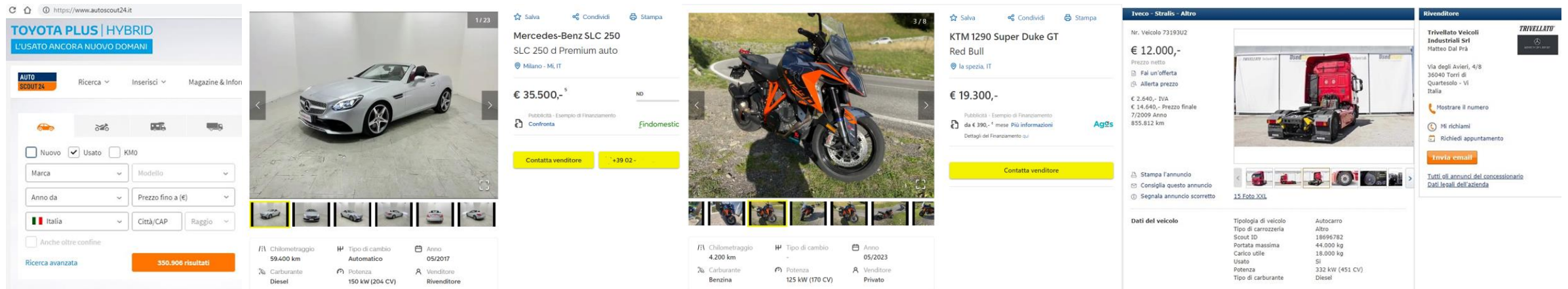
- Target del progetto
- Individuazione della fonte Big Data
- Iter procedurale per il webscraping massivo
- La tecnologia di una macro con software iMacros
- La selezione delle tipologie dei veicoli
- Architettura software del progetto
- Output prodotto dal webscraping
- La metodologia applicata per la validazione
- I volumi della fonte amministrativa di confronto
- Verifica dei risultati ottenuti
- Conclusioni

Target del progetto

- L'obiettivo è di stimare il chilometraggio medio percorso dai veicoli oggetto di annuncio di vendita per tipologia, classe di emissione, tipo alimentazione, provincia (o comune di vendita) o altre informazioni statisticamente rilevanti e confrontabili alle variabili presenti nel parco veicolare del Pubblico Registro Automobilistico (PRA) e con l'archivio delle Revisioni dei veicoli, fornito dal Ministero delle Infrastrutture e dei trasporti (MIT).
- Il calcolo dei chilometri percorsi sulla rete viaria nazionale si colloca nell'ambito di un progetto più articolato, che prevede, come risultato più ambizioso anche la stima dei flussi di traffico e di reali tassi di esposizione al rischio dell'incidentalità stradale.
- Il progetto si propone la finalità, inoltre, di effettuare comparazione e integrazione delle due tipologie di fonti Amministrativa e Big Data, per poter testare le potenzialità e la validità di entrambi gli archivi e utilizzare il valore aggiunto ottenuto dal matching delle fonti.
- Le fonti amministrative citate sono già state acquisite nel sistema di documentazione dei dati amministrativi acquisiti in Istat QRCA (Quality Report Card dei dati Amministrativi) <http://qrca.istat.it> e accessibili con specifiche autorizzazioni al trattamento dei dati.

Individuazione della fonte Big Data

- L'approccio proposto sembra essere ottimale in quanto la base dati **Autoscout24** e **TruckScout24** comprende un campione di tutti gli autoveicoli, motoveicoli e mezzi pesanti, in vendita anche nei primi 4 anni dall'immatricolazione e non soggetti quindi ancora a obblighi di revisione. Per i mezzi pesanti tale arco temporale si limita al primo anno.



The image displays three screenshots of vehicle listings from Autoscout24 and TruckScout24. The first screenshot shows a search interface for a Toyota Plus Hybrid car, with filters for 'Usato' (Used) and 'KM0'. The second screenshot shows a Mercedes-Benz SLC 250 convertible with a price of €35,500 and 59,400 km. The third screenshot shows a KTM 1290 Super Duke GT motorcycle with a price of €19,300 and 4,200 km. The fourth screenshot shows a red Iveco truck with a price of €12,000 and 855,812 km. The truck listing includes detailed specifications such as engine power (332 kW) and fuel type (Diesel).

- Nel matching per targa tra le due fonti informative per la stima del chilometraggio medio mensile percorso dai veicoli sono esclusi tutti i nuovi veicoli non ancora soggetti a revisione ed i ciclomotori.
- La fonte Big Data fornisce l'integrazione dell'informazione mancante con una stima per tali veicoli.

Iter procedurale per il webscraping massivo

I passi da seguire per la definizione dell'iter procedurale sono:

- Individuazione dei portali più attinenti alle caratteristiche della ricerca da svolgere.
- Primo contatto tramite mail con definizione degli obiettivi prefissati dallo studio.
- Alla prima risposta dal portale, si richiede il contatto telefonico con i vari soggetti responsabili del portale (responsabile marketing e responsabile IT) e successive riunioni di esemplificazione delle modalità operative.
- Predisposizioni di un documento di progetto con la definizione analitica delle modalità operative e della schedulazione del webscraping sul portale, corredato dalle finalità statistiche della ricerca e dell'indirizzo IP dal quale verranno eseguite le interrogazioni. L'inserimento dell'indirizzo IP dell'istituto nella white-list del portale risulta indispensabile per evitare blocchi all'accesso per rischi di scraping da parte di soggetti definiti dai preposti alla sicurezza informatica del portale come concorrenti.
- Invio di una lettera formale tramite PEC dell'istituto per le finalità di ricerca da svolgere.

La tecnologia di una macro con software iMacros

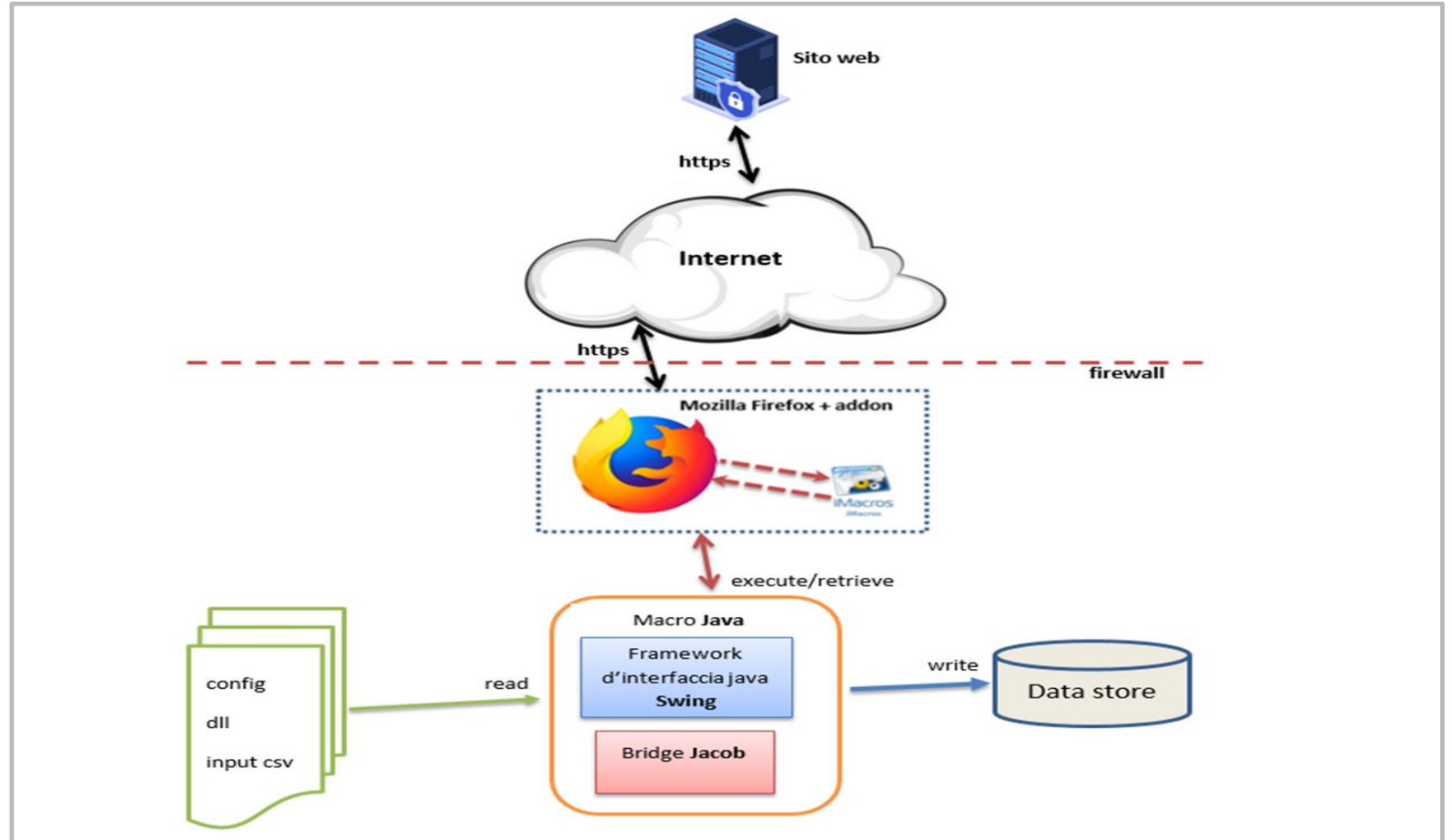
Ogni macro è realizzata in java con il software proprietario **iMacros** nella versione enterprise.

Java nella versione 8.

Il browser proprietario iMacros oppure **Mozilla Firefox** (con integrato l'addon iMacros) è utilizzato dalla macro per accedere ai siti web in https.

Ogni macro può, se necessario, interagire con l'utilizzatore grazie al framework **Swing**.

Il bridge **Jacob** consente l'interazione della macro java con il browser e l'addon.



La selezione delle tipologie dei veicoli

Le 12 macro sono state eseguite settimanalmente.

Le 6 relative a Autoscout24.it che riguardano i veicoli leggeri sono:

- Autovetture diesel revisionate (immatricolate prima del 2017) e autovetture diesel recenti (dal 2017 in poi).
- Autovetture «no diesel» (benzina, Ibride, bifuel e elettriche) revisionate , autovetture «no diesel» recenti.
- Motocicli revisionati e motocicli recenti.

Le 6 relative a Truckscout24.it che riguardano i veicoli pesanti sono:

- Autocarri, autobus, furgoni, camper, trattori stradali (motrici) e veicoli speciali.

Architettura software del progetto

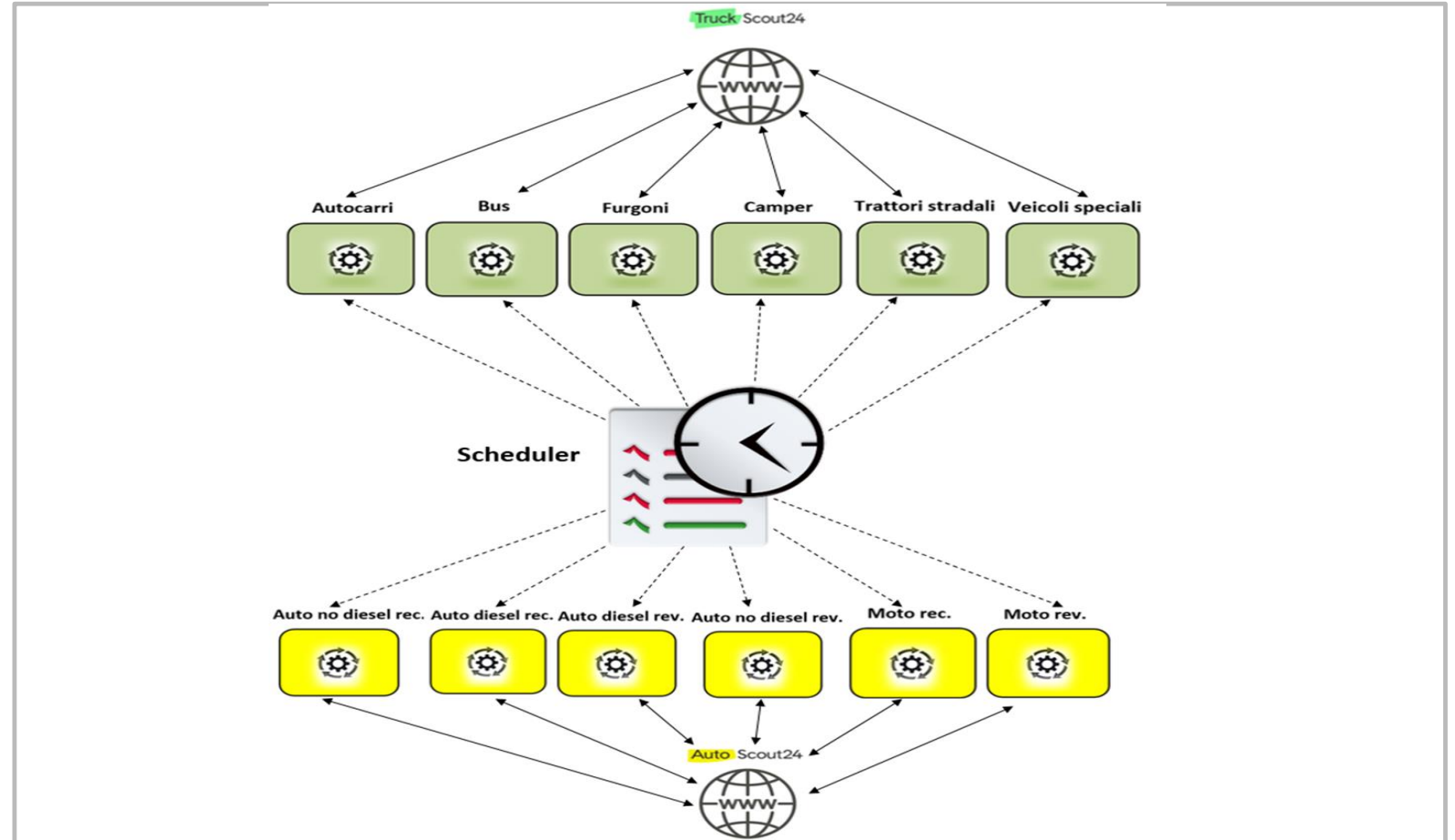
La rilevazione dei dati avviene in modo completamente automatizzato.

Lo scheduler rappresentato al centro è il programma che esegue tutte le macro secondo le tempistiche richieste.

Il webscraping su truckscout24 ed autoscout24 con due batterie di macro composta ciascuna da 6 istanze.

Gli annunci di vendita estratti sono stati:

- **319.895** per i mezzi pesanti
- **778.931** per i mezzi leggeri



Output prodotto dal webscraping

- Ciascuna macro eseguita produce un file .CSV con le informazioni catturate dalle schermate degli annunci selezionati.
- Ciascun annuncio e le corrispondenti pagine di dettaglio sono sintetizzate in una riga della matrice a fianco raffigurata come esempio.
- Tutti i file .CSV prodotti vengono poi memorizzati all'interno di un database relazionale per l'analisi statistica dei risultati.

CHILOMETRI	IMMATR.	POTENZA	SEZIONE	TIPO	SCOUT_ID	ALIM.	EMIS.	CITTA_VENDITORE	DATA_SCRITTURA	NUM_PAG.
298.500	5/1999	294 kW (400 CV)	Autocarro	Betoniera	18781907	Diesel	Euro2	25017 Lonato del Garda - E	27/03/2019 16.56	5
314.400	9/2008	332 kW (451 CV)	Autocarro	Autocarro con cassone ribaltab	18781836	Diesel	Euro4	25017 Lonato del Garda - E	27/03/2019 16.57	5
115.300	1/2013	135 kW (184 CV)	Autocarro	Pianale telonato	18781798	Diesel	Euro5	25017 Lonato del Garda - E	27/03/2019 16.57	5
422.000	9/2001	125 kW (170 CV)	Autocarro	Pianale telonato	18781678	Diesel	Euro2	25017 Lonato del Garda - E	27/03/2019 16.58	5
270.000	6/2003	290 kW (394 CV)	Autocarro	Betoniera	18781622	Diesel	Euro3	25017 Lonato del Garda - E	27/03/2019 17.00	5
1.260.200	7/2004	335 kW (455 CV)	Autocarro	Pianale telonato	18781612	Diesel	Euro3	25017 Lonato del Garda - E	27/03/2019 17.00	5
667.000	10/2005	335 kW (455 CV)	Autocarro	Autocarro con cassone ribaltab	18781600	Diesel	Euro3	25017 Lonato del Garda - E	27/03/2019 17.01	5
366.700	4/2009	353 kW (480 CV)	Autocarro	Autocarro con cassone ribaltab	18781578	Diesel	Euro5	25017 Lonato del Garda - E	27/03/2019 17.02	5
400.000	4/2003	335 kW (455 CV)	Autocarro	Autocarro con cassone ribaltab	18781287	Diesel	Euro3	29010 Pontenure - Piacenza	27/03/2019 17.03	5
222.635	12/2001		Autocarro	Pompa per calcestruzzo	18780589	Diesel	Euro3	25030 Lograto - Brescia	27/03/2019 17.05	5
554.916	2/2012	353 kW (480 CV)	Autocarro	Telaio intercambiabile	18782183	Diesel	Euro5	00155 Roma - Rm	27/03/2019 17.06	5
185.523	10/2010		Autocarro	Altro	18780463	Diesel	Euro5	31040 Pederobba - Tv	27/03/2019 17.07	5
585.000	8/1999	228 kW (310 CV)	Autocarro	Autocarro con cassone ribaltab	18779244	Diesel	Euro2	25017 Lonato del Garda - E	27/03/2019 17.08	5
280.000	7/2003	110 kW (150 CV)	Autocarro	Autoc. scarrabile	18779192	Diesel	Euro3	25017 Lonato del Garda - E	27/03/2019 17.09	5
109.900	11/2002	324 kW (441 CV)	Autocarro	Pompa per calcestruzzo	18779177	Diesel	Euro3	25017 Lonato del Garda - E	27/03/2019 17.10	5
816.000	8/2006	355 kW (483 CV)	Autocarro	Furgone silos	18779140	Diesel	Euro3	25017 Lonato del Garda - E	27/03/2019 17.10	5
495.900	6/2015		Autocarro	Autotelaio	18778800	Diesel	Euro6	70026 Modugno - Bari	27/03/2019 17.11	5
168.500	6/2015		Autocarro	Autotelaio	18778772	Diesel	Euro6	70026 Modugno - Bari	27/03/2019 17.12	5
321.412	1/2013	140 kW (190 CV)	Autocarro	Furg. trasp. refrigerato ISO	18779678	Diesel	Euro5	10024 Moncalieri - To	27/03/2019 17.15	6
765.215	3/2003	158 kW (215 CV)	Autocarro	Furg. trasp. refrigerato ISO	18779680	Diesel	Euro3	10024 Moncalieri - To	27/03/2019 17.17	6
374.533	6/2013	183 kW (249 CV)	Autocarro	Autotelaio	18779681	Diesel	Euro5	10024 Moncalieri - To	27/03/2019 17.17	6
435.793	6/2013	183 kW (249 CV)	Autocarro	Autotelaio	18779682	Diesel	Euro5	10024 Moncalieri - To	27/03/2019 17.18	6
390.000	6/2003		Autocarro	Autoc. scarrabile	18776177	Diesel	Euro3	29010 Pontenure - Piacenza	27/03/2019 17.26	6
297.000	8/2012	183 kW (249 CV)	Autocarro	Altro	18775588	Diesel	Euro4	10024 Moncalieri - To	27/03/2019 17.27	6
464.000	6/2007		Autocarro	Furg. trasp. refrigerato ISO	17924955	Diesel	Euro4	20090 Cusago - Mi	27/03/2019 17.27	6
450.000	2/2007		Autocarro	Pianale telonato	18695162	Diesel	Euro4	80024 Cardito - Na	27/03/2019 17.28	6
326.432	6/2005		Autocarro	Altro	18775686	Diesel	Euro2	36016 Thiene - Vi	27/03/2019 17.29	6

La metodologia applicata per la validazione

- L'introduzione della targa dei veicoli nella fornitura dell'archivio del Parco Veicolare dal 2020, ha permesso di agganciare la corrispondente chiave di associazione con i dati della fonte informativa delle revisioni veicoli, ove risulta presente il totale dei chilometri percorsi e le relative date di riferimento (immatricolazione e revisione) per ciascuna epoca nel periodo 2014-2021. La normativa vigente nella revisione dei veicoli, biennale per i mezzi leggeri e dopo i primi quattro anni per la prima revisione implica l'analisi su un periodo superiore e cronologicamente precedente all'anno di rilevazione del fenomeno. Per i mezzi pesanti la revisione annuale permette meglio di verificare la congruenza delle informazioni elaborate tramite il join dei due database di fonte amministrativa con caratteristiche di totalità del fenomeno.
- La comparazione tra i due metodi avviene sul valore esaustivo per i veicoli che hanno eseguito almeno una prima revisione ovvero immatricolati prima del 2017.
- Per coloro che non sono stati soggetti ancora di revisione, l'integrazione della fonte Big Data risulta essere una stima attendibile in quanto non presente nessuna informazione relativa al chilometraggio totale effettuato.

I volumi della fonte amministrativa di confronto

Parco veicolare al 2020	Meno di 2 anni	Tra i 2 a 5 anni	Tra i 5 a 10 anni	Tra i 10 a 20 anni	Oltre i 20 anni	Totale
Autoveicoli per trasporto di persone						
AB - AUTOBUS	6.936	15.377	14.863	40.177	22.530	99.883
AC - AUTOCARAVAN	13.110	20.438	28.403	126.574	109.864	298.389
AP - AUTOVEICOLO PER TRASPORTO PROMISCUO	1	14	52	7.328	439.690	447.085
AV - AUTOVETTURE	3.526.322	7.451.583	7.263.123	14.765.581	6.264.180	39.270.789
Autoveicoli per trasporto di merci						
AM - AUTOCARRO	317.884	662.193	564.894	1.531.516	1.145.231	4.221.718
AS - AUTOVEICOLO SPECIALE	33.477	72.957	69.635	162.308	127.971	466.348
TS - TRATTORE STRADALE	21.744	53.184	33.665	61.543	25.333	195.469
Motoveicoli						
MC - MOTOCICLO	444.684	761.540	937.214	2.809.591	2.031.293	6.984.322
MM - MOTOCARRO	638	3.184	6.306	29.690	176.953	216.771
MP - MOTOVEICOLO PROMISCUO			2	16	66	84
MZ - MOTOCICLO CON CARROZZETTA	3.810	4.525	1.965	638	8.274	19.212
QC - QUADRICICLO	2.943	10.805	21.115	70.613	9.973	115.449
Veicoli trainati e non classificati						
RM - RIMORCHIO	24.172	61.696	40.487	138.750	149.693	414.798
Non classificato			1	4	18	23
Totale	4.395.721	9.117.496	8.981.725	19.744.329	10.511.069	52.750.340

Revisioni	2014	2015	2016	2017	2018	2019	2020	Totale
Autoveicoli	4.200.525	10.985.000	10.890.578	13.580.702	14.443.370	14.877.732	13.795.060	82.772.967
Motoveicoli	187.219	1.013.243	1.095.757	1.417.395	1.480.989	1.522.451	1.307.941	8.024.995
Totale	4.387.744	11.998.243	11.986.335	14.998.097	15.924.359	16.400.183	15.103.001	90.797.962

Verifica dei risultati ottenuti

- L'analisi del chilometraggio medio annuo per i veicoli leggeri mostrano distribuzioni simili tra le due fonti.
- Da evidenziare che dai criteri di analisi da fonte amministrativa, il metodo di calcolo del chilometraggio medio per i veicoli meno recenti per effetto delle successive revisioni biennali, rendono più basso il valore medio riscontrato rispetto al valore medio annuo tra la data di immatricolazione e quella di vendita.

Media Km percorsi	Anno di immatricolazione						
Autovetture	Recenti	Prima revisione	Tra i 5 a 10 anni	Tra i 10 a 20 anni	Oltre i 20 anni	Media	Totale veicoli
Fonte amministrativa		15.728	13.515	11.328	6.802	12.724	34.485.667
Fonte Big Data	15.662	17.110	15.646	11.689	4.925	14.870	476.045
Alimentazione Gasolio	Recenti	Prima revisione	Tra i 5 a 10 anni	Tra i 10 a 20 anni	Oltre i 20 anni	Media	Totale veicoli
Fonte amministrativa		18.798	15.846	13.804	11.972	15.938	15.913.065
Fonte Big Data	16.389	20.191	17.589	14.195	8.550	17.867	116.196
Media Km percorsi	Anno di immatricolazione						
Motocicli	Recenti	Prima revisione	Tra i 5 a 10 anni	Tra i 10 a 20 anni	Oltre i 20 anni	Media	Totale veicoli
Fonte amministrativa		4.118	3.349	2.712	1.911	3.064	4.653.892
Fonte Big Data	5.371	4.307	3.768	2.910	1.584	3.601	226.458

Conclusioni

- Il target del progetto è la costruzione di modelli di stima dei chilometri percorsi dai veicoli sul territorio nazionale al fine della definizione degli indicatori di “veicoli teorici medi giornalieri” (V.T.M.G.) che rappresentano il numero di veicoli che ogni giorno percorrono l'intera rete stradale dato dal rapporto fra i chilometri percorsi sul tratto e la sua lunghezza in chilometri moltiplicata per il numero dei giorni. Questo misura il grado di utilizzo della rete, dell'autostrada o della tratta considerata. Al termine del processo sarà possibile rappresentare una stima più aderente all'esposizione di rischio degli utenti della strada per le diverse categorie di veicoli.
- Insieme agli indicatori di incidentalità stradale basati sull'estesa chilometrica, elemento di novità rispetto alla tradizionale produzione, su popolazione residente o parco veicolare, consentono di eliminare parte degli effetti della componente di mobilità del fenomeno e di effettuare un confronto territoriale più adeguato, congiuntamente ad altri progetti di localizzazione mirata dei flussi di traffico sui singoli territori, tipologie di strade o singoli tratti, rappresenta l'obiettivo d'individuazione puntuale di esposizione al rischio di essere coinvolto in incidente stradale per le diverse categorie di utenti.
- Inoltre con la profilazione dei veicoli coinvolti in incidenti stradali, aspetto mai considerato nelle analisi sull'incidentalità stradale, mediante il matching con le targhe dei veicoli si potrà categorizzare i veicoli coinvolti in base alle loro percorrenze medie.



Laboratorio innovazione.

Risultati e prospettive

24
NOVEMBRE
2023

GRAZIE
PER L'ATTENZIONE

[#IstatWebinar](#)