

8 GIUGNO 2018

Stime delle modalità di utilizzo dei siti web da parte delle imprese

L'Istat rende disponibili stime dell'utilizzo dei siti web da parte delle imprese, prodotte utilizzando direttamente quanto disponibile in Internet. Tali stime riportano il tasso di imprese sul totale della popolazione di riferimento che possiede o utilizza un sito web attraverso il quale:

1. offrono funzioni di web ordering (componente dell'e-commerce);
2. propongono offerte di lavoro;
3. espongono link a social media (Facebook, Twitter, Instagram ecc.).

Le stime, riferite al 2017, riguardano una popolazione di riferimento che nel suo complesso include circa 184.000 imprese con più di 10 addetti operanti in diversi settori di attività economica e vengono calcolate, per ogni variabile di interesse, oltre che a livello nazionale, anche per i seguenti domini:

- le 4 classi di addetti;
- le 16 modalità risultanti dall'incrocio tra classi di addetti e i quattro macrosettori di attività economica;
- le 26 classi di attività economica;
- le 21 regioni e province autonome.

Tali stime sono correntemente prodotte con tecniche tradizionali (raccolta dati mediante questionario web presso un campione di imprese) ogni anno dall'indagine Istat, comune a tutti i paesi membri della UE, sull'uso delle tecnologie dell'informazione e della comunicazione (ICT) da parte delle imprese¹.

¹ La rilevazione annuale sulle tecnologie dell'informazione e della comunicazione (ICT) viene effettuata dal 2001 dall'Istat per produrre indicatori sull'uso dell'ICT nelle imprese con almeno 10 addetti attive nell'industria e nei servizi non finanziari. Dal 2004 l'indagine fa parte delle statistiche comunitarie sulla società dell'informazione previste dal regolamento della Commissione n. 808/2004 in base al quale i contenuti del questionario vengono armonizzati e resi comuni tra tutti gli Stati Membri. L'indagine fornisce informazioni sulle attività svolte dall'impresa su Internet (sito web, cloud computing, social media), tipologie di connessione utilizzate (banda larga fissa e mobile), sull'utilizzo di strumenti informatici adottati per gestire l'attività interna (ad esempio utilizzo di software ERP, CRM) o i rapporti con i referenti della filiera produttiva (clienti, fornitori, altre imprese), sul commercio elettronico, l'utilizzo di competenze specialistiche in ICT, la fatturazione elettronica. L'indagine ICT è una delle principali fonti di dati per il Digital Agenda Scoreboard che viene utilizzato dalla Commissione europea per misurare il progresso dell'economia digitale europea e contribuisce con circa 11 indicatori a definire l'Indice di digitalizzazione dell'economia e della società (DESI).

Le stime sperimentali qui riportate sono ottenute mediante una procedura alternativa che si articola nelle seguenti fasi:

1. gli indirizzi dei relativi siti web (circa 100.000) sono stati reperiti in parte attraverso una procedura che ha fatto uso dei motori di ricerca e delle informazioni anagrafiche contenute nell'Archivio Statistico delle Imprese Attive (ASIA), in parte utilizzando informazioni già disponibili (da indagine o dall'archivio CONSODATA);
2. mediante gli indirizzi ottenuti, si è proceduto a effettuare web scraping dei siti e a catturarne dove possibile il contenuto testuale;
3. i circa 85.000 testi così ottenuti sono stati elaborati con tecniche di Natural Language Processing al fine di ottenere termini (n-grams) rilevanti ai fini della predizione delle diverse variabili target;
4. nel sottoinsieme di unità per le quali erano disponibili sia le risposte fornite all'indagine che i testi catturati nel web (poco meno di 12.000) è stato possibile mettere a punto degli strumenti di predizione del valore (presenza/assenza) di ognuna delle tre variabili target sulla base dei termini rilevanti individuati al passo precedente; nel caso del web ordering e dell'offerta online di lavoro sono stati utilizzati modelli "random forest", mentre nel caso dei social media, tecniche di "information retrieval";
5. tali strumenti sono stati applicati alle 85.000 imprese per cui era stato possibile raccogliere ed elaborare i testi contenuti all'interno dei corrispondenti siti;
6. le stime alternative rispetto a quelle correnti dell'indagine (stime da indagine) sono state calcolate con due stimatori diversi: uno full model based (stime da modello) e l'altro basato su una combinazione dei valori predetti e dei dati di indagine (stime combinate).

Nei file excel resi disponibili (uno per ogni variabile di interesse) sono riportati:

- il particolare dominio di classificazione;
- la numerosità campionaria dell'indagine sull'uso dell'ICT da parte delle imprese;
- la stima della variabile di interesse correntemente prodotta dall'indagine, assieme agli estremi del relativo intervallo di confidenza;
- la stima da modello basata sul conteggio dei valori predetti per ogni singolo sito acceduto;
- la stima combinata, facente uso cioè sia dei dati di indagine che dei valori predetti.

Inoltre, viene presentata una tabella solo con le stime da modello per le tre variabili di interesse per attività economica a livello di due cifre Ateco (62 divisioni). Tale dettaglio rappresenta nuova informazione rispetto a quella prodotta dall'indagine che può fornire solo un dettaglio settoriale più aggregato.

Con riferimento ai fenomeni stimati è opportuno precisare alcune differenze tra quelli definiti dal questionario e dal [Manuale metodologico di indagine](#) e quelli osservati da modello.

Nell'anno 2017 l'indagine infatti chiede alle imprese con almeno 10 addetti che dichiarano di avere una connessione a Internet, se esse hanno un sito web e, al sottoinsieme che risponde positivamente, vengono richieste quali funzionalità o servizi sono disponibili sul sito stesso. Il quesito non si riferisce esclusivamente al sito web proprio dell'impresa o di sua proprietà, ma all'utilizzo di un sito web da parte dell'impresa comprendendo, quindi, anche siti web di terzi come, ad esempio, il sito web del gruppo di imprese di cui fa parte l'impresa rispondente, il sito web di una impresa holding o affiliata a quella rispondente che comunque abbia pagine Internet dedicate all'impresa rispondente².

Per quanto riguarda la funzione di web ordering offerta dal sito web ovvero la possibilità di effettuare ordinazioni o prenotazioni online, i rispondenti sono invitati a non considerare come presente tale funzionalità se, per concludere l'operazione, vi è la necessità di contatti offline supplementari o via email mentre il pagamento può non essere incluso nella procedura. Relativamente alla presenza sul sito web di un link ai social media, si fa riferimento alla presenza di link a un account o profilo dell'impresa rispondente su uno specifico social media. Infine, nel 2017, solo nel questionario italiano e non anche in quello europeo, si richiede se sul sito web vengono pubblicati annunci di posti di lavoro vacanti o se viene data la possibilità di effettuare domande di impiego on-line, comprendendo sia la presenza di semplici informazioni sulle offerte di lavoro, sia la presenza di strumenti on-line per presentare candidature.

Alla produzione di queste stime hanno collaborato:

- Istat: G. Barcaroli, G. Bianchi, N. Golini, A. Nurra, P. Righi, S. Salamone, F. Scalfati, M. Scannapieco, D. Summa, D. Zardetto
- CINECA: M. Scarnò
- Università di Roma Sapienza: R. Bruni

² Non sono incluse tra quelle che hanno un sito web e imprese presenti solo in siti di *e-marketplace* (come Amazon, Booking, Expedia, Groupon, ecc.) nei quali hanno la possibilità di pubblicizzare e vendere i loro prezzi o quelle che compaiono in siti web di terzi solo con il nome, l'indirizzo, informazioni di contatto (ad es. Pagine Gialle).