

What do we gain from ML?

Autocoding real estate webscraped data

Dominik Dąbrowski

D.Dabrowski@stat.gov.pl

Why Machine Learning?

Seeking ways to augment
official statistics







How we started our work

- Created a division for organising and producing experimental analysis
- ML project - Automatic classification of online shops products

Download data

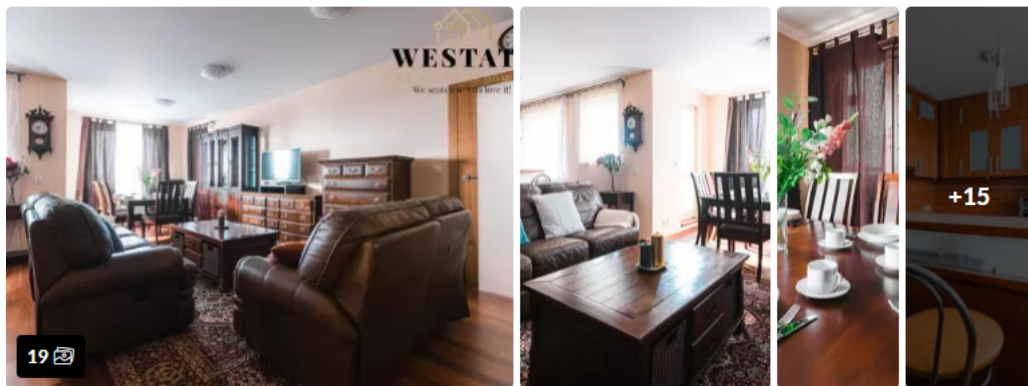


Warszawa, Mokotów, mazowieckie



Mieszkanie Warszawa Mokotów, ul. Bukowińska

986 000 zł 13 683,04 zł/m² REKLAMA [Sprawdź ratę](#)



Szczegóły ogłoszenia

Lokalizacja Warszawa, Mokotów, mazowieckie	Powierzchnia w m2 72,06 m ²	Liczba pokoi 3
Piętro 9	Stan wysoki standard	Rok budowy 2006
Typ zabudowy blok	Materiał budynku cegła	Liczba miejsc parkingowych 1
Czy mieszkanie ma łazienkę? tak	Numer referencyjny gratka-531/10239/OMS	Zaktualizowane dziś

Opis ogłoszenia

ENG BELOW
TYLKO U NAS!!!

Westate prezentuje przestronne mieszkanie na Mokotowie na sprzedaż.

Ta nieruchomość to 72,06m² powierzchni użytkowej + 6,93m² loggia

W skład nieruchomości wchodzi:

- sypialnia z łazienką (z wanną)
- salon z otwartą kuchnią i wyjściem na zabudowaną loggię
- łazienka (z prysznicem)
- gabinet

W budynku znajduje się basen i siłownia dla mieszkańców z osobnym wejściem.

Do mieszkania przynależy miejsce postojowe w garażu podziemnym.

W okolicy liczna infrastruktura usługowa nie będąca uciążliwa dla mieszkańców.

Celem uzyskania większej ilości informacji zapraszamy do kontaktu:

Westate Wilanów Sp. Zo.o.

ul. Branickiego 15,

02-972 Warszawa

tel: **+48...** [pokaż numer](#)

	Price: 5100 PLN/m ²	Level: 3	furnished: YES	market: secondary
type: block of flats	Area: 50 m ²	No. of rooms: 2		

[Finanse: Oblicz ratę dla tego mieszkania >>](#)

elevator, intercom



Opis

Sprzedam ładne słoneczne mieszkanie na os. Uroczysko ul. Warszawska piętro 3 /winda, domofon czysta klatka blok docieplony po termomodernizacji, spokojni sąsiedzi tylko 3 mieszkania na pięttrze .

Mieszkanie bez wad! Od pierwszego właściciela mieszkania !

Mieszkanie jest z lat 80 tych ,słoneczne dwustronne środkowe.

Mieszkanie ma bardzo ładny rozkład

Pokój 19,80m² z balkonem → balcony

Pokój 10m² m.ay

Kuchnia 8,5 oddzielna z oknem → separate kitchen with window

Łazienka 3,5m

Garderoba 5m → separate wardrobe

przedpokój 7m

Pomieszczenia są ustawne w kwadracie

Garderoba 5m dwie i pół szafy w przedpokoju szeroka wnęka

w mieszkaniu były wymienione okna plastikowe Jezierski nowe 6 tys niemieckie profile

Tynki zrobione, ściany są równe,okna nowe mieszkanie jest ciepłe są zwroty za ogrzewanie 600zł ,niski czynsz 300zł

Super punkt bardzo fajne mieszkanie o pięknym rozkładzie można uzyskać 3 pokoje przez przeniesienie kuchni do pokoju który ma 20 m uzyskując aneks kuchenny +2 oddzielne sypialnie!

Wolne od zaraz szybka sprzedaż zapraszam

cena 255tys z piwnica w cenie czyli gratis → basement

Prepare basic cleaning

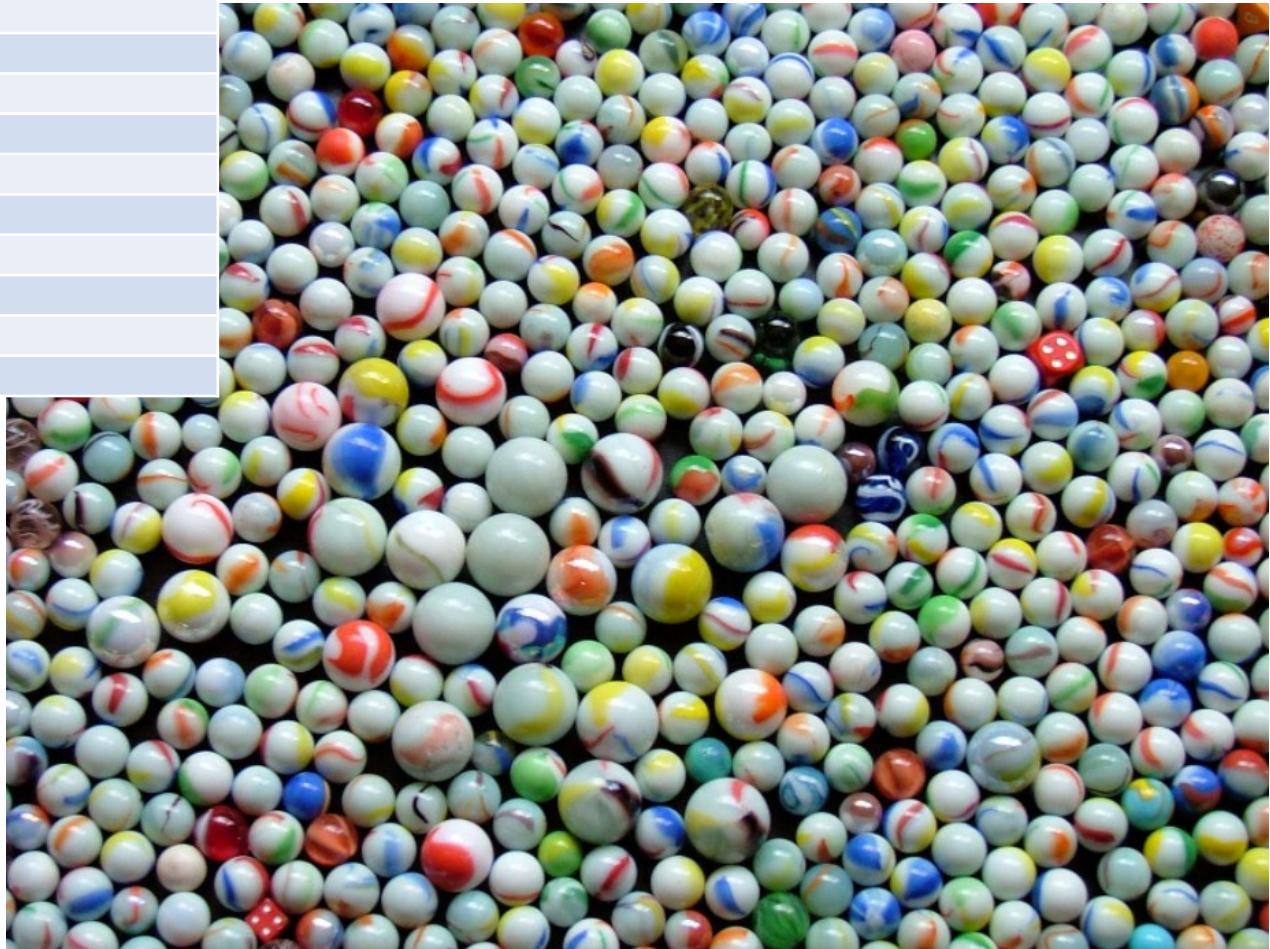


Problems we encountered

- Unstructured webscraped data
- Long text variables with multiple various information
- Duplicates

Not homogeneous data (patchy data)

Length	Number of words
94	7
110	14
128	16
134	14
151	18
...	...
9797	1312
9797	1312
12842	1779
12923	1787
14141	1995



Offers of multiple objects

Multifunctional apartments – perfect investment in Poznan, Downtown. Apartments are located in 10-storey building on first floor. Total area: 372 m²

It consists of:

1. studio apartment 19.12 m²
 2. studio apartment 18.34 m²
 3. studio apartment 18.51 m²
 4. studio apartment 19.18 m²
 5. studio apartment 19.91 m²
 6. studio apartment 19.78 m²
 7. studio apartment 19.19 m²
 8. 2-room apartment 34.59 m²
 9. 2-room apartment 37.05 m²
 10. studio apartment 19.66 m²
 11. studio apartment 18.15 m²
 12. studio apartment 18.59 m²
 13. 2-room apartment 25.17 m²
 14. studio apartment with balcony 20.5 m²
- and common corridor with an area 64 m².

Duplicated descriptions

Description	Price	Currency	Area	Address
KUPNO OD DEWELOPERA, 0% prowizji, PLANOWANE TERMINY ZAKOŃCZENIA INWESTYCJI - budynek A - IV kwartał 2021, budynek B - II kwartał 2022Budynek A lokal: 1.A.8.13, balkon [Mieszkania z opcją SMART HOME, pakiet antysmogowy]Cena brutto m. p. w garażu podziemnym: - 35 000 zł (naziemne - 18 000 zł) - ILOŚĆ OGRANICZONA !!	351241	PLN	34.52	Hetmańska, Poznań
KUPNO OD DEWELOPERA, 0% prowizji, PLANOWANE TERMINY ZAKOŃCZENIA INWESTYCJI - budynek A - IV kwartał 2021, budynek B - II kwartał 2022Budynek A lokal: 1.A.8.11, balkon [Mieszkania z opcją SMART HOME, pakiet antysmogowy]Cena brutto m. p. w garażu podziemnym: - 35 000 zł (naziemne - 18 000 zł) - ILOŚĆ OGRANICZONA !!	447526	PLN	48.91	Hetmańska, Poznań
KUPNO OD DEWELOPERA, 0% prowizji, PLANOWANE TERMINY ZAKOŃCZENIA INWESTYCJI - budynek A - IV kwartał 2021, budynek B - II kwartał 2022Budynek A lokal: 1.A.8.6, balkon [Mieszkania z opcją SMART HOME, pakiet antysmogowy]Cena brutto m. p. w garażu podziemnym: - 35 000 zł (naziemne - 18 000 zł) - ILOŚĆ OGRANICZONA !!	459450	PLN	51.05	Hetmańska, Poznań

Achieve the goal



Why ML methods over traditional approaches?

- Less errors/mistakes?

Why ML methods over traditional approaches?

- Less errors/mistakes?
- More time efficient

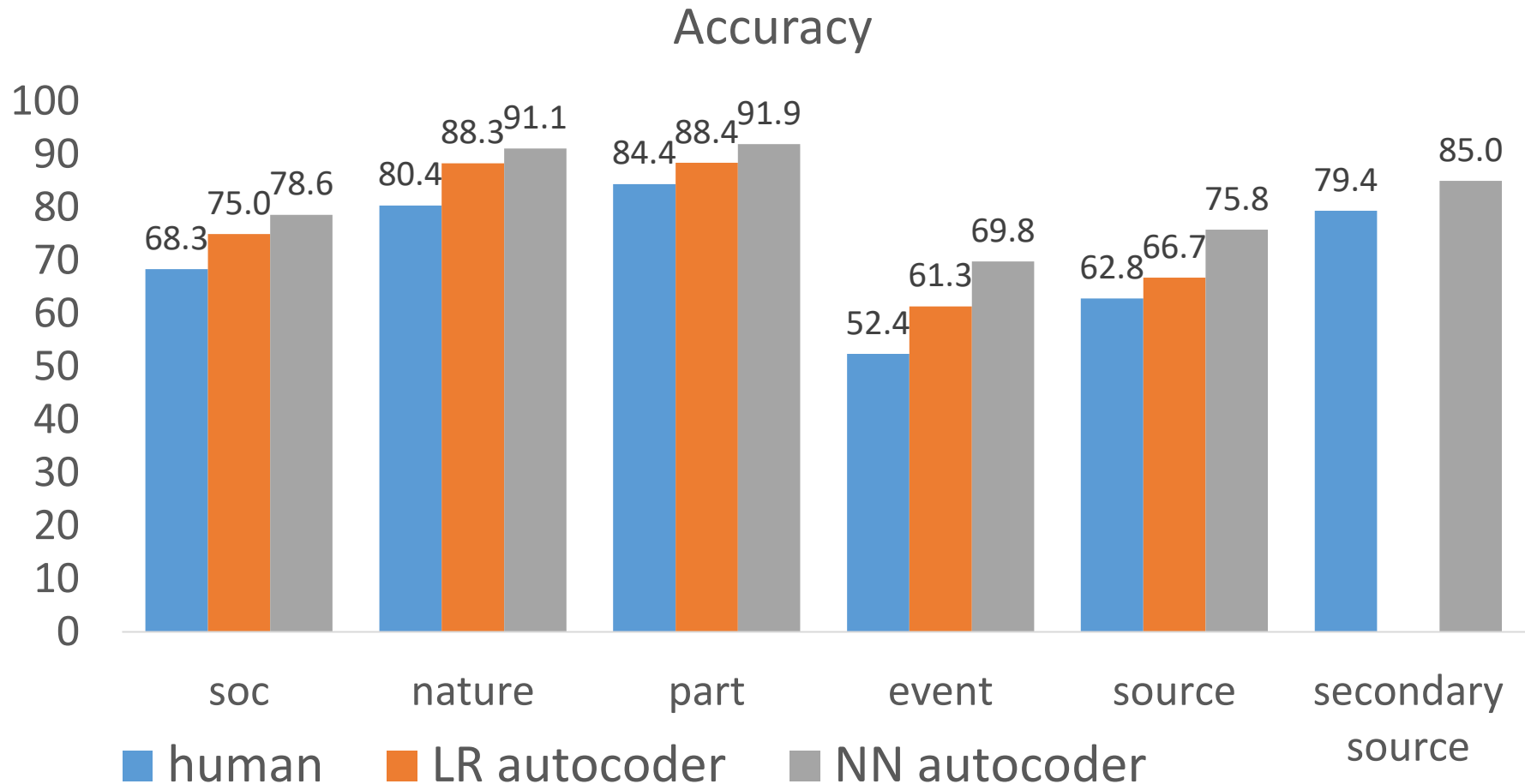
Why ML methods over traditional approaches?

- Less errors/mistakes?
- More time efficient
- Replicability

Our results on autocoding

- More than 72% of the automatically classified data with probability over 95%.
- 4 people manually classified 6688 from 8881 non-duplicated offers (1239, 2565, 1750, 1134)

BLS data comparison of methods



Autocoding the Survey of Occupational Injuries and Illnesses – 5 years in, U.S. Bureau of Labor Statistics, Alexander Measure

What we gain from the experiment

Learn

What we gain from the experiment

Learn

Get involved

What we gain from the experiment

Learn

Get involved

Share

Thank you

For more information contact:

Dominik Dabrowski

D.Dabrowski@stat.gov.pl