Tasks, resources and evaluation PART-2

Linguistic Resources for Natural Language Processing LM Language Technologies and Digital Humanities 2024-25

Cristina Bosco

Beyond accuracy

Another metric, with respect to accuracy, for better evaluating system performance is based on the **separate observation** the performance of the model with respect to the different **categories to be classified** for the task.

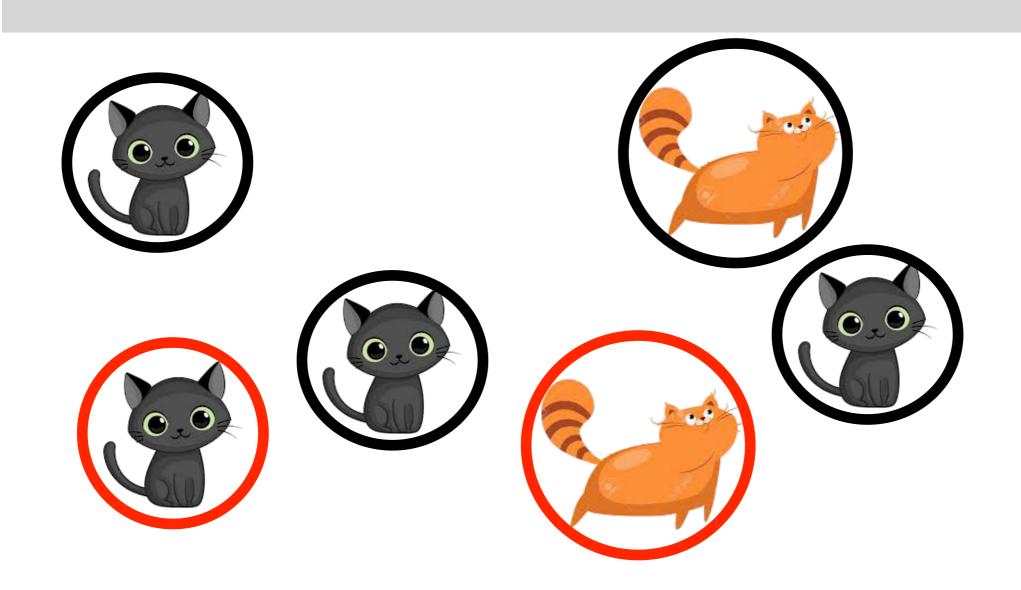
We focus in turn on a single category only, by not considering the other ones and evaluating the system against it alone.

Some statistical notions



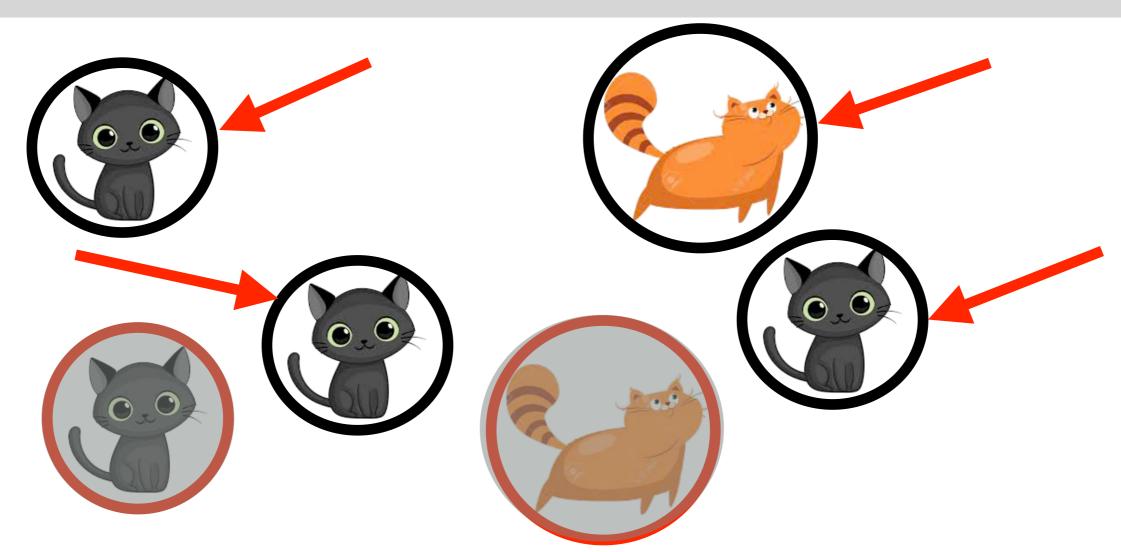
In a **gold standard** set of cats, each element is labeled as **BLACK** or as **RED** (the picture shows how they are labeled in the gold).

Some statistical notions



A model (whose output is not perfect, as usually!) labels the set of cats as **BLACK** or as **RED** (the circles shows how the model labels each of them).

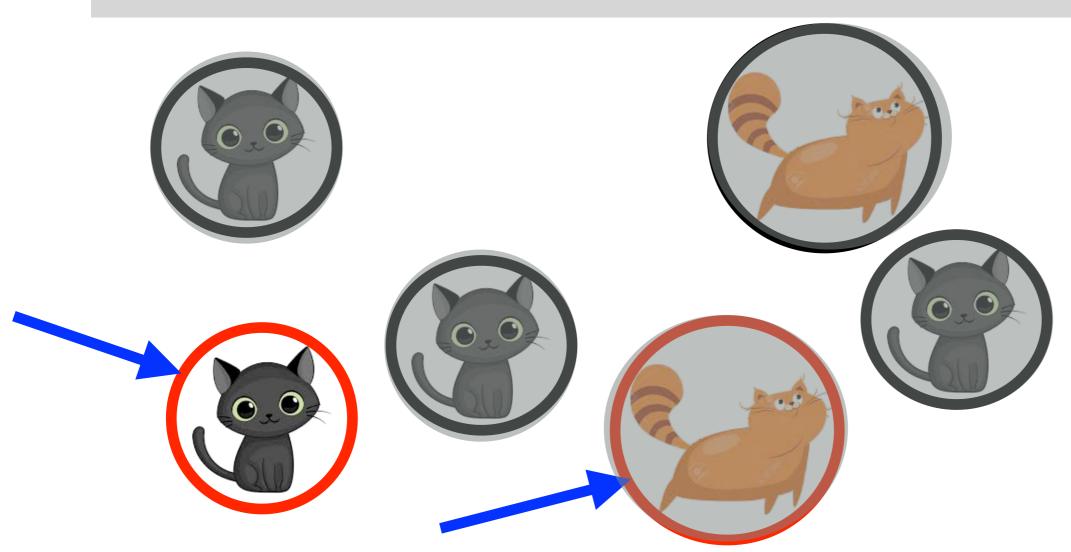
Positives and Negatives



If we focus on the category Black only:

- all the cats the model labels as Black are called **POSITIVE** instances of Black (indicated by red arrows)

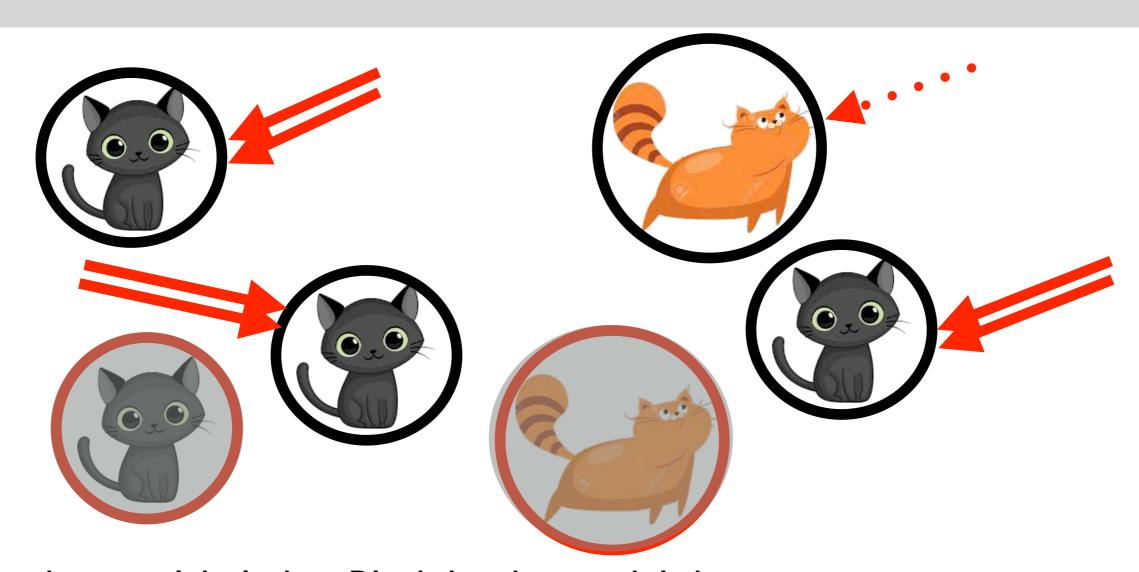
Positives and Negatives



If we focus on the category Black only:

- all the cats the model labels as not Black (in this case Red) are called **NEGATIVE** instances of Black (indicated by blue arrows)

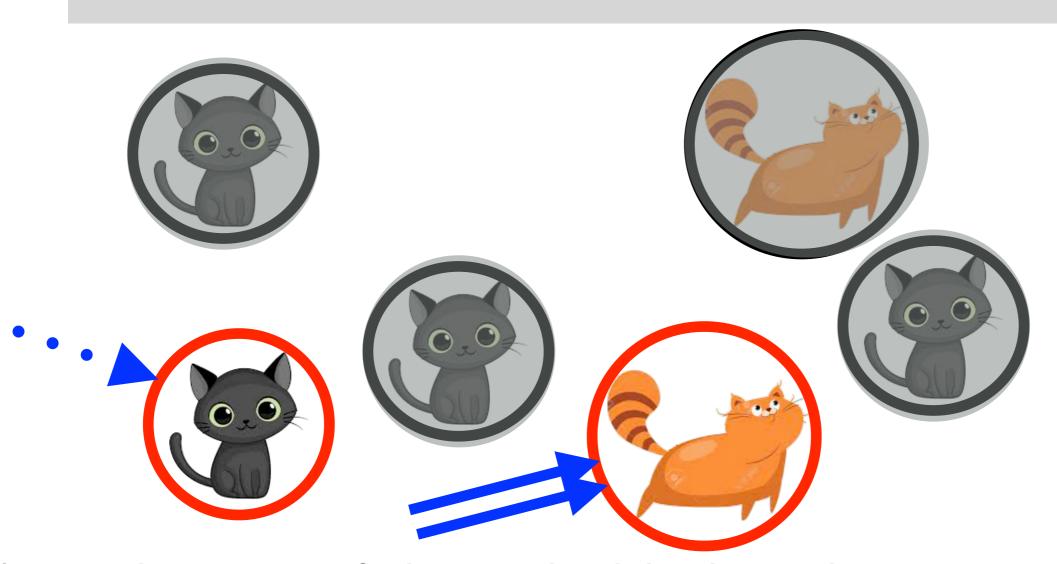
True and False instances



Among the cats labeled as Black by the model there are:

- TRUEPOSITIVEs that the model labels as Black and the gold standard too (indicated by double red arrow)
- a **FALSEPOSITIVE** that the model labels as Black but the gold standard does not (indicated by a dashed red arrow)

True and False instances



Among the cats classified as non-black by the model are:

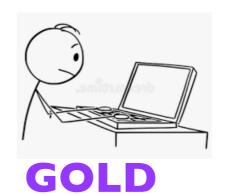
- a **TRUENEGATIVE**, which the model classifies as other than Black and also the gold standard (indicated by double blue arrow)
- a **FALSENEGATIVE**, which the model classifies as other than Black, but the gold standard does not (indicated by a dashed blue arrow)

Partial PoS tagging

(with 2 tags only)

TEST SET

Model output



nonVERB

nonVERB

nonVERB

nonVERB

nonVERB

VERB

The	VERB
cat	nonVERB
run	nonVERB
in	nonVERB
the	nonVERB
garden	VERB

Mary sleeps in the sun	nonVERB VERB nonVERB nonVERB VERB	nonVERB VERB nonVERB nonVERB nonVERB
sun	VERB	nonVERB

Two categories are annotated in the gold standard test set and must be annotated by the model. They are VERB and nonVERB and can be separately observed in the evaluation process.

True and false positives and negatives

Separately observing the class nonVERB to be categorised by the model, we can provide more precise evaluations of its performance. The model classifies:

- as nonVERB cat, in, the, Mary, in, the, that are nonVERB in the gold: they are the TRUE POSITIVES

Partial PoS tagging (with 2 tags only)

TEST SET

Model output



TRUE nonVERB positives

The cat run in the garden	VERB nonVERB nonVERB nonVERB VERB	nonVERB NonVERB nonVERB nonVERB nonVERB	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Mary sleeps in the sun	nonVERB VERB nonVERB nonVERB VERB	nonVERB VERB nonVERB nonVERB nonVERB	

The model outputs 6 TRUE POSITIVEs (TPs) for the class nonVERB

= tokens the model classifies as nonVERB and the gold too

Partial PoS tagging (with 2 tags only)





TRUE nonVERB positives

TEST SET

Model output

The	
cat	
run	
in	
the	
garden	

VERB	nonVERB
nonVERB	nonVERB
nonVERB	VERB
nonVERB	nonVERB
nonVERB	nonVERB
VERB	nonVERB

0	
0	
0	

Mary	
sleeps	
in	
the	
sun	

nonVERB	nonVERB
VERB	VERB
nonVERB	nonVERB
nonVERB	nonVERB
VERB	nonVERB
V LIND	

The model outputs 6 TRUE POSITIVEs (TPs) for the class nonVERB

= tokens the model classifies as nonVERB and the gold too

True and false positives and negatives

Separately observing the class nonVERB to be categorised by the model, we can provide more precise evaluations of its performance. The model classifies:

- as nonVERB cat, in, the, Mary, in, the, that are nonVERB in the gold: they are the TRUE POSITIVES
- as nonVERB run, that is VERB in the gold: it is a **FALSE POSITIVE**

Partial PoS tagging (with 2 tags only)

GOLD

FALSE nonVERB positives

The	VERB	nonVERB	0
cat	nonVERB	nonVERB	l
run	nonVERB	VERB	0
in	nonVERB	nonVERB	I
the	nonVERB	nonVERB	
garden	VERB	nonVERB	0
Mary	nonVERB	nonVERB	I
sleeps	VERB	VERB	0
in	nonVERB	nonVERB	I
the	nonVERB	nonVERB	1
sun	VERB	nonVERB	0

The model outputs I FALSE POSITIVE (FP) for the class nonVERB

= a token the model classifies as nonVERB but the gold classifies otherwise

(with 2 tags only) **nonVERB** positives **TEST SET Model output** GOLD The **VERB** nonVERB nonVERB nonVERB cat nonVERB **VERB** run nonVERB nonVERB in nonVERB the nonVERB nonVERB garden **VERB** nonVERB nonVERB Mary **VERB** sleeps **VERB** nonVERB nonVERB in nonVERB nonVERB the **VERB** nonVERB sun

FALSE

Partial PoS tagging

The model outputs I FALSE POSITIVE (FP) for the class nonVERB

= tokens the model classifies as nonVERB but the gold classifies otherwise

True and false positives and negatives

Separately observing the class nonVERB to be categorised by the model, we can provide more precise evaluations of its performance. The model classifies:

- as nonVERB cat, in, the, Mary, in, the, that are nonVERB in the gold: they are the TRUE POSITIVES
- as nonVERB run, that is VERB in the gold: it is a **FALSE POSITIVE**
- as VERB sleeps, that is VERB in the gold: it is a TRUE NEGATIVE

Partial PoS tagging (with 2 tags only)

TEST SET	TES'	T S	ET	
----------	------	-----	----	--

Model output



TRUE nonVERB negatives

The	VERB	nonVERB	0
cat	nonVERB	nonVERB	
run	nonVERB	VERB	0
in	nonVERB	nonVERB	
the	nonVERB	nonVERB	
garden	VERB	nonVERB	0
Mary	nonVERB	nonVERB	I
sleeps	VERB	VERB	0
in	nonVERB	nonVERB	
the	nonVERB	nonVERB	
sun	VERB	nonVERB	0
		• • •	• • •

The model outputs I TRUE NEGATIVE (TN) for the class nonVERB

= tokens the model classifies as VERB and the gold too

Partial PoS tagging (with 2 tags only)

	7
COLD	

TRUE nonVERB negatives

																											TEST SET	Model outpu	ıt
												i Lu i u i i u u ci u ci u ci u ci u ci		LS I SL I I'IOGEI OGLDGI	LS I SL I MOUEL OULDUL	LS I SL I MUUEI UULDUL	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output								
I ES I SE I MOUEL OULDUL	I ES I SE I MOUEL OULDUL	I ES I SE I MOUEL OULDUL	I ES I SE I MOUELOULDUL	LES I SE I MOUEL OULDUL	legi sei Mudel uuluul												TEST SET Model quitout	TEST SET Model quitout	TEST SET Model quitaut	TEST SET Model quitaut	TEST SET Model quitaut	TEST SET Model quitaut	TEST SET Model quitaut						
i E3 i 3E i i i i i ouei outbut	i E3 i 3E i i i i i ouei outbut	i E3 i 3E i i i i i ouei outbut	I ES I SE I MOUELOULDUL	I ES I SE I MOUELOULDUL		LES SE MOUELOULDUL											TEST SET Model quitout	TEST SET Model quitout	TEST SET Model quitout	TEST SET Model quitaut	TECT CET Madal authort	TECT CET Madal authout	TECT CET Model authort	TECT CET Model quitaut	TEST SET Model quitaut	TEST SET Model quitout			
i E3 i 3E i Model outbut	i E3 i 3E i Model outbut	i E3 i 3E i Model outbut	i es i se i model outbut	les i se i model outbut	les i se i Model outbut												TEST SET Model quitout	TEST SET Model quitout	TEST SET Model quitout	TECT CET Madal authort	TECT CET Madal authort	TECT CET Madal authout	TECT CET Madal authort	TECT CET Model authort	TECT CET Model authors	TECT CET Model quenue			
I ES I SE I MOGEI OULDUL	I ES I SE I MOGEI OULDUL	I ES I SE I MOGEI OULDUL	I ES I SE I MOGEI OULDUL	i es i se i Model outbut	LES I SE I MOUELOULDUL	I ES I SE I MOUELOULDUL											TECT CET Model entenut	TECT CET Model entenut	TECT CET Model entenut	TEST SET Model output	TECT CET Madal authort	TECT CET Madal authort	TECT CET Madal authort	TECT CET Model authors	TECT CET Model authors	TECT CET Model authors			
i es i se i Model outbut	i es i se i Model outbut	i es i se i Model outbut	i es i se i Model outbut	i es i se i model outbut	les i se i model outbut	les i se i Model Outbut											TECT CET Model entent	TECT CET Madal autout	TECT CET Madal authors	TECT CET Model entenut	TECT CET Madal authors	TECT CET Madal authorit	TECT CET Madal authors	TECT CET Madal authors	TECT CET Madal authors	TECT CET Model authors			
i es i se i Model outbut	i es i se i Model outbut	i es i se i Model outbut	i es i se i Model outbut	i es i se i model outbut	leg i se i model ourbut	les i se i model outbut	rs se model outbut										TECT CET Madalautuut	TECT CET Madalautaut	TECT CET Madalautuut	TECT CET Madalautuut	TECT CET Madalauteur	TECT CET Madalauteur	TECT CET Madalauteur	TECT CET Madalauteut	TECT CET Madalauteut	TECT CET Model autout			
I E2 I 2E I Model outbut	I E2 I 2E I Model outbut	I E2 I 2E I Model outbut	i co i oci model outbut	i co i oci model outbut	i es i se i Model outbut	i co i oci model outbut	LEST SET MODEL OUTDUT	TO THE MANAGEMENT OF THE PROPERTY OF THE PROPE			I PO I SPI IVIANDI ALITALIT	ΙΕΝΙΙΚΟΙΙ ΙΝΙΛΟΙΙ ΑΙΙΤΟΙΙΤ	· · · · · · · · · · · · · · · · · · ·	I - S I S I I I I I I I I I I I I I I I			TECT CET Madalautusut	TECT CET Madalaution	TECT CET Madalauteur	TECT CET Madalauteut	TECT CET Madalaukuuk	TECT CET Madalaukuuk	TECT CET Madalauteur	TECT CET Madalautaut	TECT CET Madalautaut	TECT CET Madalautaut			
i co i octout	i co i octout	i co i octout	i co i octout	les i se i model outbut	LES SE MODEL OUTDUT	LES SE MOGEL OUTDUT	Model outbut	TO THE MODEL OUTDUT	TO MODEL OUTDUT	TO MODEL OUTDUT	ΙΕΝΙΙΚΟΙ ΜΙΛΛΟΙ ΛΙΙΤΝΙΙΤ	WINNEL MINNEL	I I I I I I I I I I I I I I I I I I I	THE STANDAL ALITALITY	I I I I I I I I I I I I I I I I I I I											TECT CET Madalanda.			
I ES I SE I Model outbut	I ES I SE I Model outbut	I ES I SE I Model outbut	E3 3E Model output	i co i octout	LES I SE I MOGELOUTDUT	LES SE MOGEL OUTDUT	Model Output	Model output	MODEL MICHEL	MODEL MODEL OUTDUT	PIONE MINNER ALITHUT	MODEL MINDER OUTDIT		MINADI MITALIT		WINANI ALITALIT										TECT CET Madalanda.			
E3 3E Model output	E3 3E Model output	E3 3E Model output	E3 3E Model output	i es i se i model outbut	LES SE Model output	Model output	Model output	Model Output	MODEL MICHEL	MODEL MICHEL	MANA ALITALIT	MINAGI ALITALIT	MANA ALITALIT	MINABI MINABI MITALIT	MINADI MITALIT	NIAAAI MIAAAI AIITAIIT													
E5 5E Model output	E5 5E Model output	E5 5E Model output	les i se i model output	les i se i model output	LES SE Model output	LES SE MODEL OUTDUT	Model output	Model Output	MODEL OUTDUT	MODEL MINDE	MOGOL OUTDUT	MANAL MITALIT	MAAAI AIITAIIT		MAAAI AIITAIIT	MODOL OUTDUT													
LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	LES I SE I MODEL OUTDUT	MOGEL OUTDUT	Model output	Model output	MODEL OUTDUT	MAGAI AIITHIIT	MAGAI AUTHUT	MOGOL OUTDUT	MAGAI AUTHUT	MAGAI AUTHUT	MODOL OUTDUIT													
LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	MODEL OUTDUT	Model output	Model output	MODEL OUTDUT	MAGAI AUTHUT	MAGAI AIITHIIT	MOGOLOUTHUIT	MOGOL OUTDUIT	MAGGI AUTHUT	MAGAI AUTHUT													
LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	LES SE Model output	LES I SE I Model output	LEST SET MODEL OUTDUT	Model output	Model output	Model output	MODEL MICHAEL	MODEL MICHAEL	MAGAI AUTHUT	MAGAI AUTHUT	MODEL MODEL OUT DUT	MAGAI AUTHUR													
TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	Lest Set Model output	Model output	Model output	Model output	Model output	Model cuthut	MODE OUTDUIT	MAGAI AUTHUT	MAGAI AUTHUT	THE MAGAI AUTHUR	The Madal Author													
TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model quitnut	TEST SET Model quithlif	TEST SET Model quithlif	TEST SET Model quithlif	TEST SET Model quithlit	TEST SET Model quithlit	TEST SET Model quitaut													
TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model outnut	TEST SET Model quithlif	TEST SET MADA Authur	TEST SET Model quithlit	TEST SET Model quithlit	TEST SET Model quithlit	TEST SET Model quithlit	TECT SET Model cuthout													
TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model outnut	TEST SET Model quithuit	TEST SET Model quithuit	TEST SET Model quithuit	TEST SET Model quithlit	TEST SET Model quithuit	TEST SET Model quithilit													
TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model outnut	TEST SET Model quithuit	TEST SET Model quithut	TEST SET Model quithut	TEST SET Model quithlit	TEST SET Model quithuit	TEST SET Model quithiut													
TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model quitnut	TEST SET Model quithut	TEST SET Model quithut	TEST SET Model quithut	TECT CET Model quithut	TEST SET Model output	TEST SET Model quithiut													
TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model quitaut													
TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output													
TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output	TEST SET Model output													

The	VERB	nonVERB	0
cat	nonVERB	nonVERB	I
run	nonVERB	VERB	0
in	nonVERB	nonVERB	I
the	nonVERB	nonVERB	I
garden	VERB	nonVERB	0

Mary	nonVERB	nonVERB	I
sleeps	VERB	VERB	0 <
in	nonVERB	nonVERB	I
the	nonVERB	nonVERB	I
sun	VERB	nonVERB	0

The model outputs I TRUE NEGATIVE (TN) for the class nonVERB

= tokens the model classifies as VERB and the gold too

True and false positives and negatives

Separately observing the class nonVERB to be categorised by the model, we can provide more precise evaluations of its performance. The model classifies:

- as nonVERB cat, in, the, Mary, in, the, that are nonVERB in the gold: they are the TRUE POSITIVES
- as nonVERB run, that is VERB in the gold: it is a **FALSE POSITIVE**
- as VERB sleeps, that is VERB in the gold: it is a TRUE NEGATIVE
- as VERB the, garden, sun, that are nonVERB in the gold: they are the FALSE NEGATIVEs.

Partial PoS tagging (with 2 tags only)

negatives **TEST SET Model output** GOLD The **VERB** nonVERB nonVERB nonVERB cat nonVERB **VERB** run nonVERB nonVERB in the nonVERB nonVERB nonVERB garden **VERB** nonVERB nonVERB Mary **VERB** sleeps **VERB** nonVERB nonVERB in nonVERB nonVERB the **VERB** nonVERB sun

FALSE

nonVERB

The model outputs 3 FALSE NEGATIVE (Fns) for the class nonVERB

= tokens the model classifies otherwise but the gold classifies as nonVERB

(with 2 tags only) nonVERB negatives **TEST SET Model output** GOLD **VERB** nonVERB The nonVERB nonVERB cat nonVERB **VERB** run nonVERB nonVERB in the nonVERB nonVERB **VERB** nonVERB garden nonVERB nonVERB Mary **VERB** sleeps **VERB** nonVERB nonVERB in nonVERB nonVERB the **VERB** nonVERB sun

FALSE

Partial PoS tagging

The model outputs 3 FALSE NEGATIVE (Fns) for the class nonVERB

= tokens the model classifies otherwise but the gold classifies as nonVERB

Dividing the number of TRUE POSITIVEs classified by the model for the total of all the tokens it classifies (as TRUE POSITIVEs or as FALSE POSITIVEs) we can calculate a measure called PRECISION.

Precision answers the question:

Are the tokens classified as nonVERB in effect nonVERB (in the gold)?

Partial PoS tagging

(with 2 tags only)

TEST SET

Model output



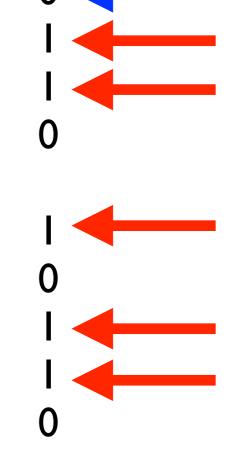
TRUE and FALSE positives

The	
cat	
run	
in	
the	
garden	

VERB
nonVERB
nonVERB
nonVERB
nonVERB
VERB
nonVERB

nonVERE nonVERE VERB nonVERE nonVERE
nonVERE VERB nonVERE nonVERE
nonvekt

nonVERB



Mary sleeps in the sun nonVERB VERB nonVERB nonVERB VERB

Given 6 TRUEPositives and I FALSEPositive,

the Precision of the model is 6:(6+1)=0.85

Maximum precision = 1:

a model that classifies only nonVERB tokens with category nonVERB; all classifications provided by the model are correct and correspond to the gold; the model provides only TRUE POSITIVES and no FALSE POSITIVES.

TRUEpositives

FALSEpositives

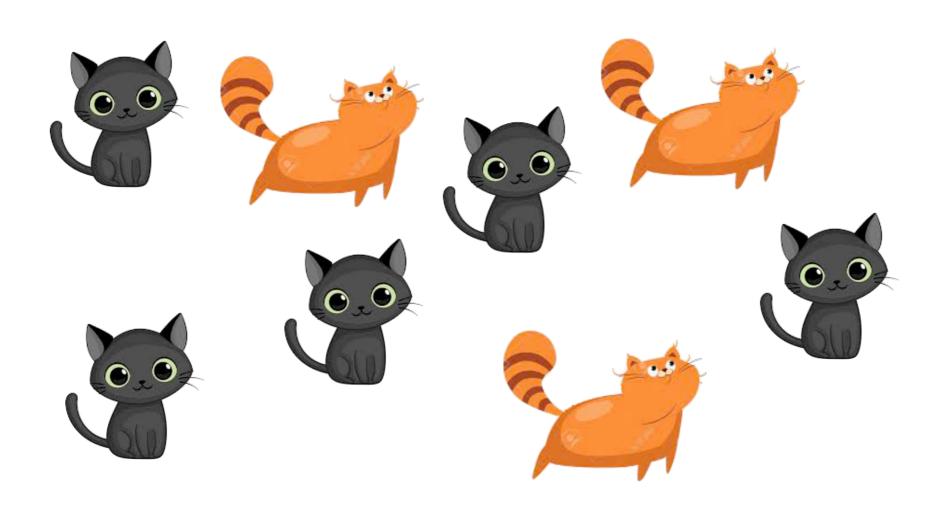
Minimum precision = 0:

a model that does not classify any nonVERB token with category nonVERB; all classifications provided by the model do not correspond to the gold; the model provides only FALSE POSITIVES and no TRUE POSITIVES.

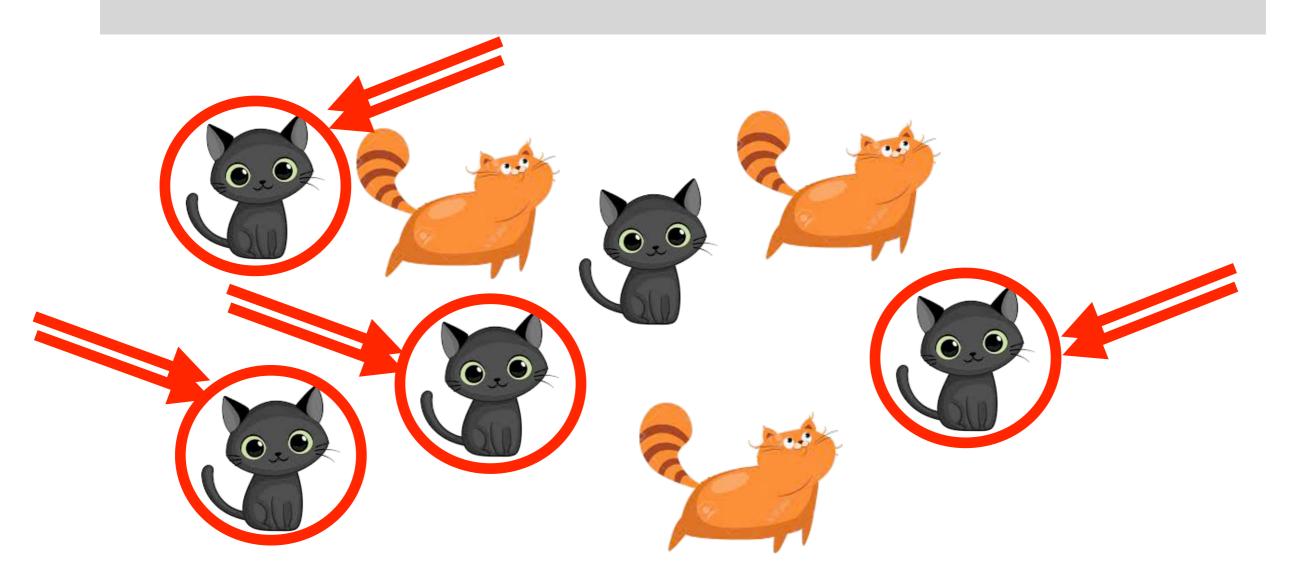
Is precision sufficient to evaluate the result of the model?

Precision measure tells us only if the classified tokens have been classified correctly.

But Precision does not say anything about the amount of tokens classified by the model.

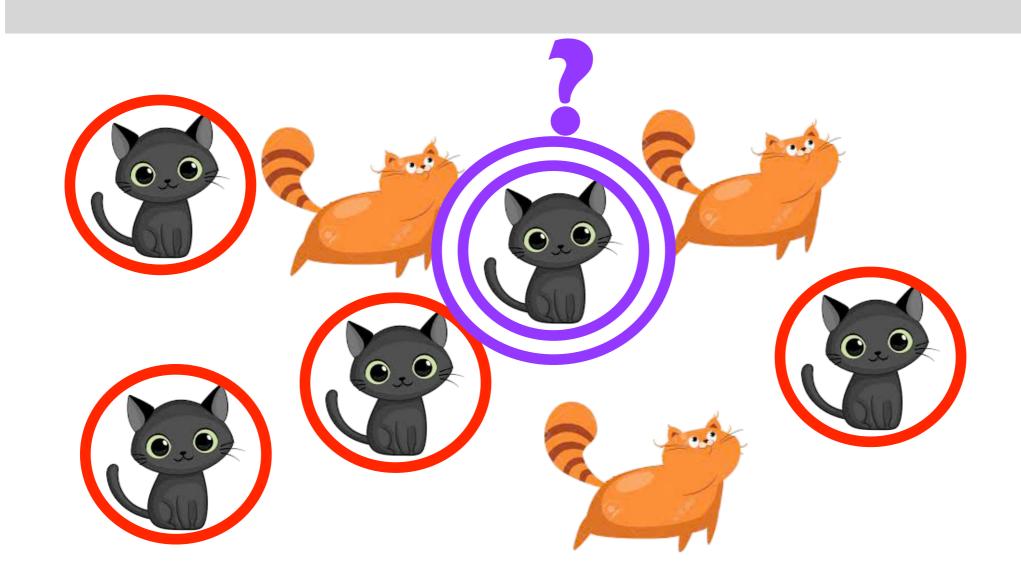


The model has to classify the category black cats in the set above.



The system identifies 4 black cats. Its precision is 4 TruePositives: (4 TruePositives + 0 FalsePositives) = I

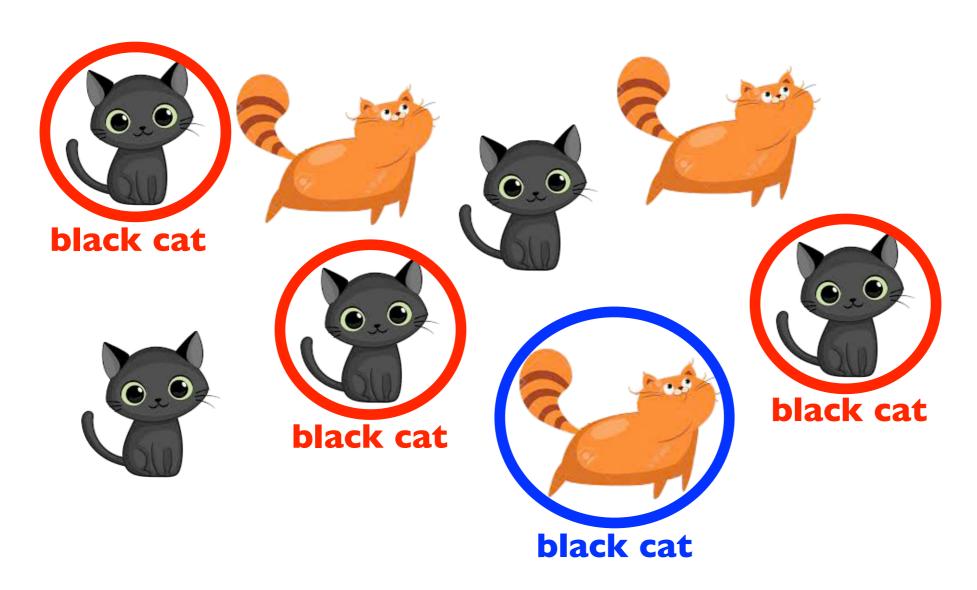
A very good result! But is this model really good?



Although the model reaches the precision maximum score (I), it fails to identify some examples of the class to be identified.



The precision maximum score can be also achieved by models that fail to identify several examples ... if they do not generate **FALSEpositives!**



The precision score decreases only when the model classifies as black cats also some red cats:

3 TruePositives: (3 TruePositives + | FalsePositive) = 0.75

Precision and Recall

Precision alone is not enough!

Precision measure tells us only if the classified tokens have been classified correctly. It is only sensitive to misclassification.

Another measure called RECALL tell us how many of the tokens that must be classified are actually classified.

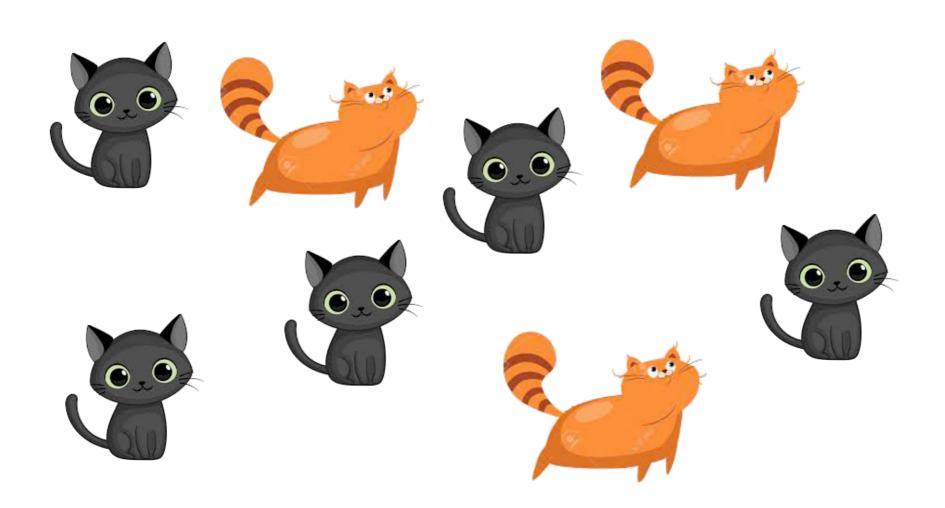
Dividing the number of TRUE POSITIVEs classified by the model for the total of all the tokens it classifies (as TRUE POSITIVEs or as FALSE NEGATIVEs) we can calculate a measure called RECALL.

Recall answers the question:

How many of the tokens that must be classified with a given category are actually classified by the model with that category?

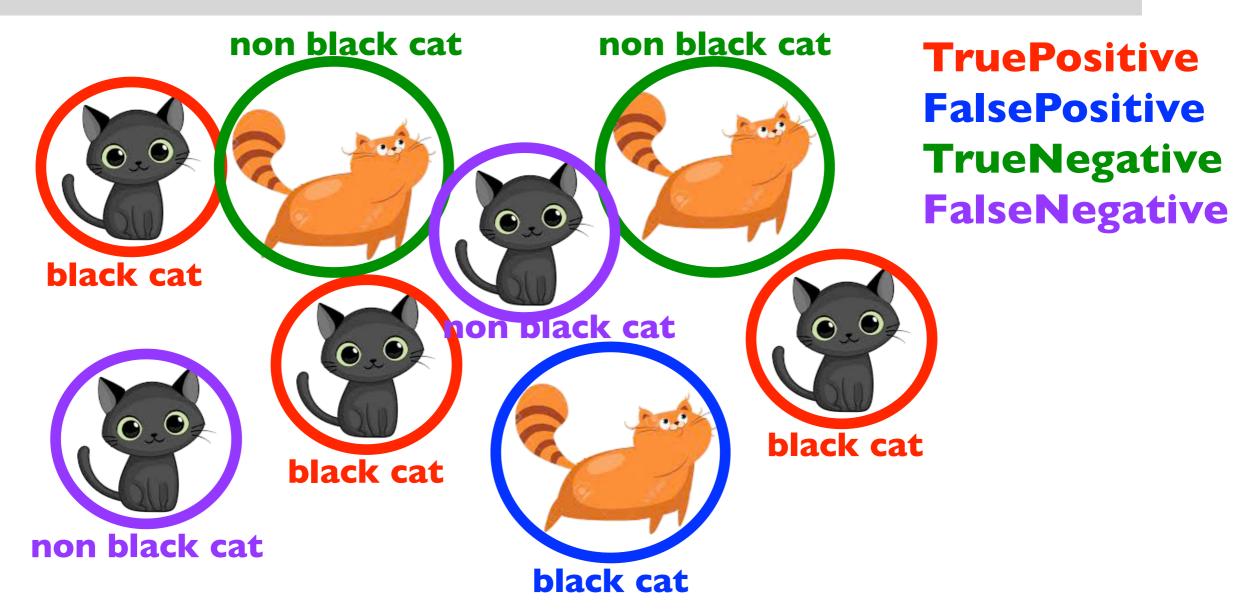
The sum of TruePositives and FalseNegatives for a given class exactly corresponds to the number of tokens that are categorised with that class in the gold.

Among them TruePositives are those correctly classified by the model, while FalseNegatives are those uncorrectly classified by the model.

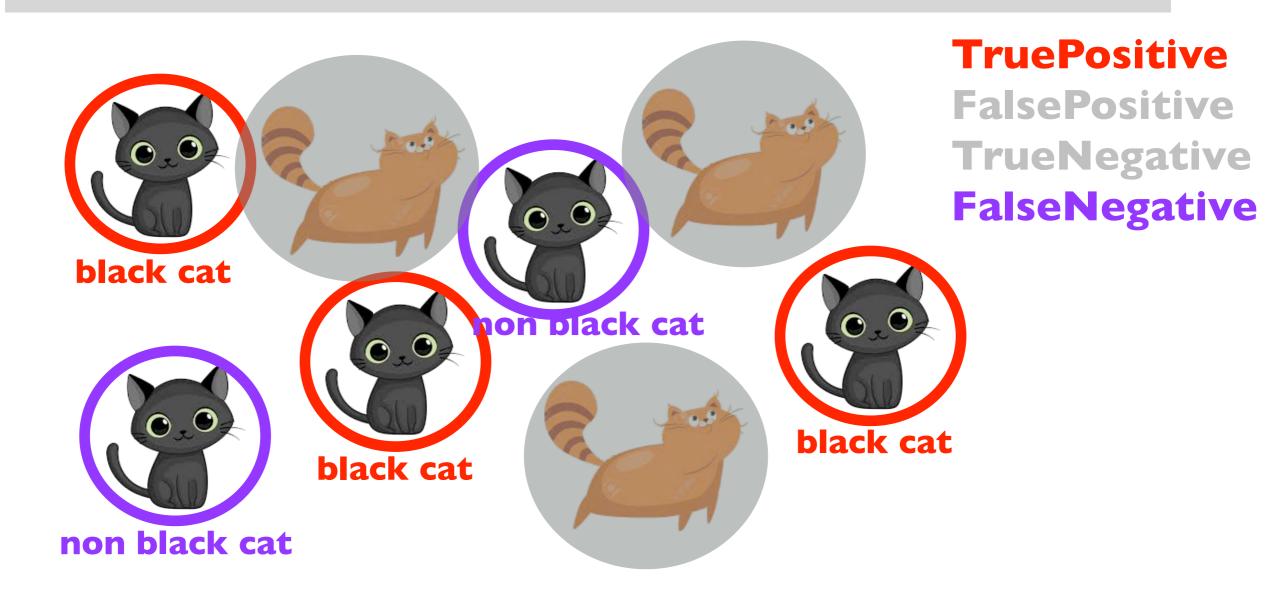


The model has to classify the category black cats in the set above.

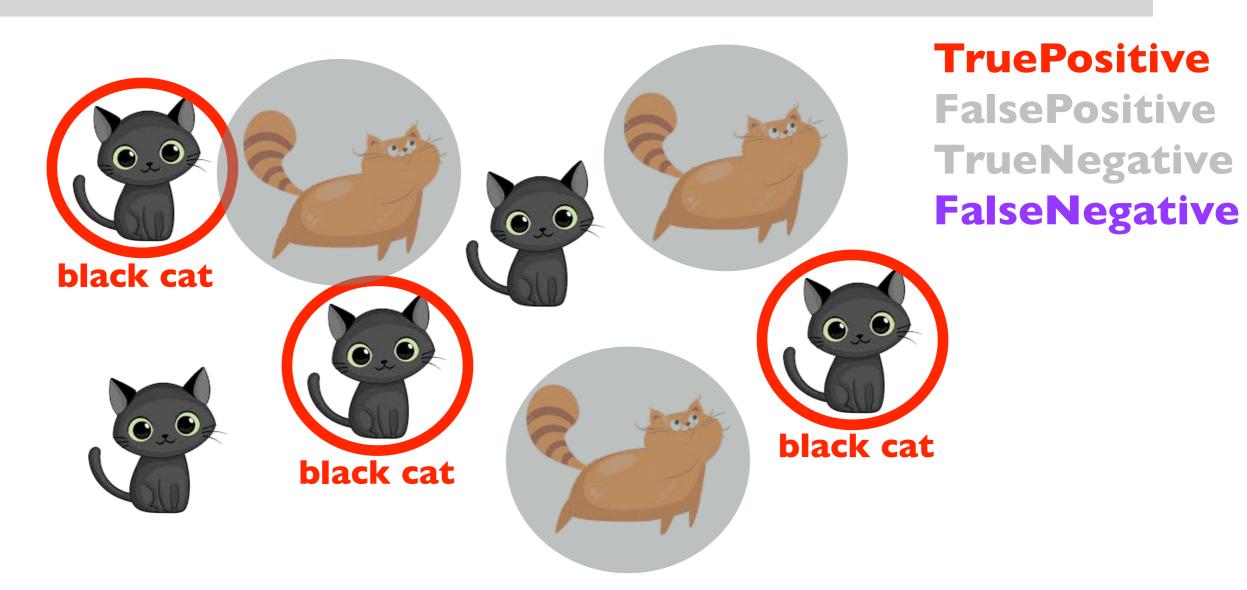
We see that the elements classified as black cats in the gold are 5.



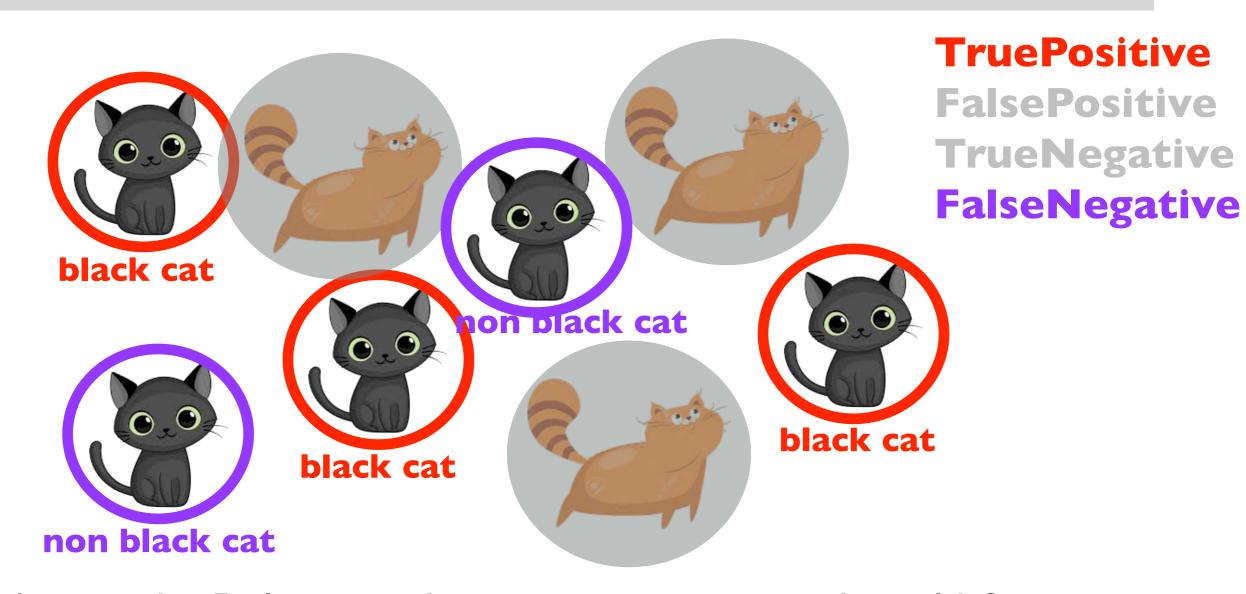
The model classifies the category **black cat**: some elements are correctly classified (TPs) and some incorrectly (FPs). It classifies also the category **non black cat** generating some TNs and some Fns.



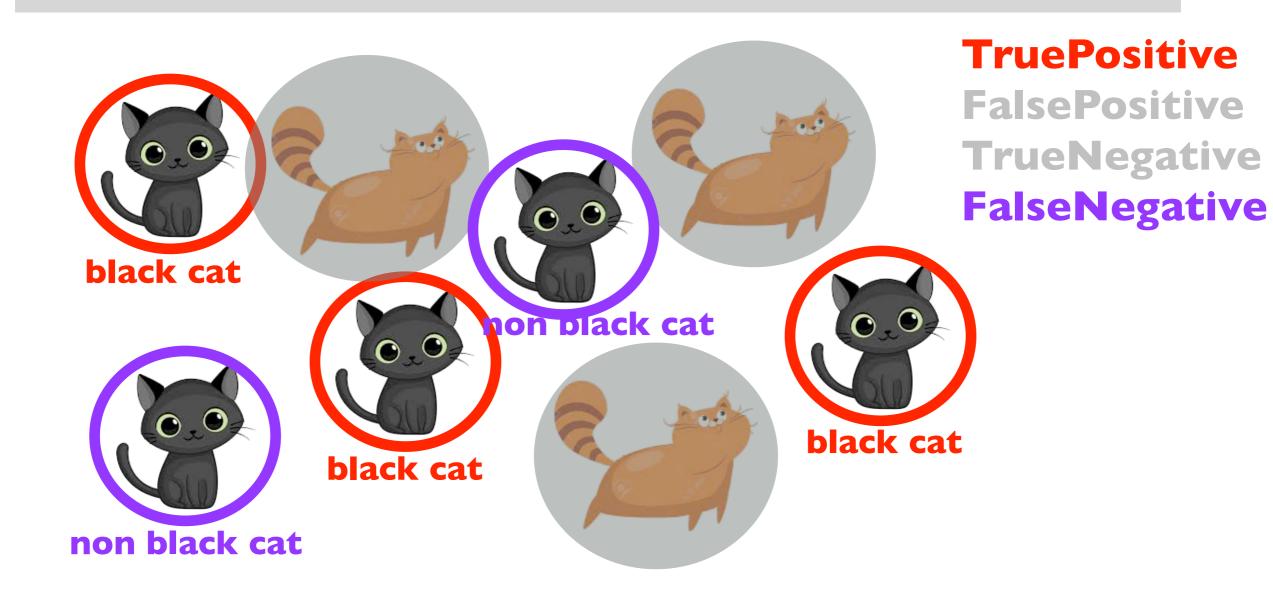
We want to focus only on the elements that are classified as black cat in the gold.



Among the 5 elements that are black cats in the gold, 3 are classified by the model as black cats (TPs)



Among the 5 elements that are black cats in the gold, 3 are classified by the model as black cats (TPs) while 2 are classified by the model as non black cats (FNs).



The system identifies 3 black cats (over 5). Its recall is

3 TruePositives: (3 TruePositives + 2 FalseNegatives) = 0.6

Partial PoS tagging

(with 2 tags only)

TEST SET

Model output



TRUE positives and FALSE negatives

The	VERB
cat	nonVERB
run	nonVERB
in	nonVERB
the	nonVERB
garden	VERB

ın	nonVEKB
the	nonVERB
garden	VERB
Mary	nonVERB
sleeps	VERB
in	nonVERB
the	nonVERB
sun	VERB
• • •	• • •

nonVERB	0
nonVERB	
VERB	0
nonVERB	
nonVERB	
nonVERB	0
nonVERB VERB	0
. —	
nonVERB	
nonVERB	
nonVERB	0

Given 6 TRUEPositives and 3 FALSENegatives,

the Recall of the model is 6:(6+3)=0.66

Maximum recall = 1:

a model that classifies all nonVERB tokens with category nonVERB; all classifications provided by the model are correct and correspond to the gold; the model provides only TRUE POSITIVES and no FALSE NEGATIVES.

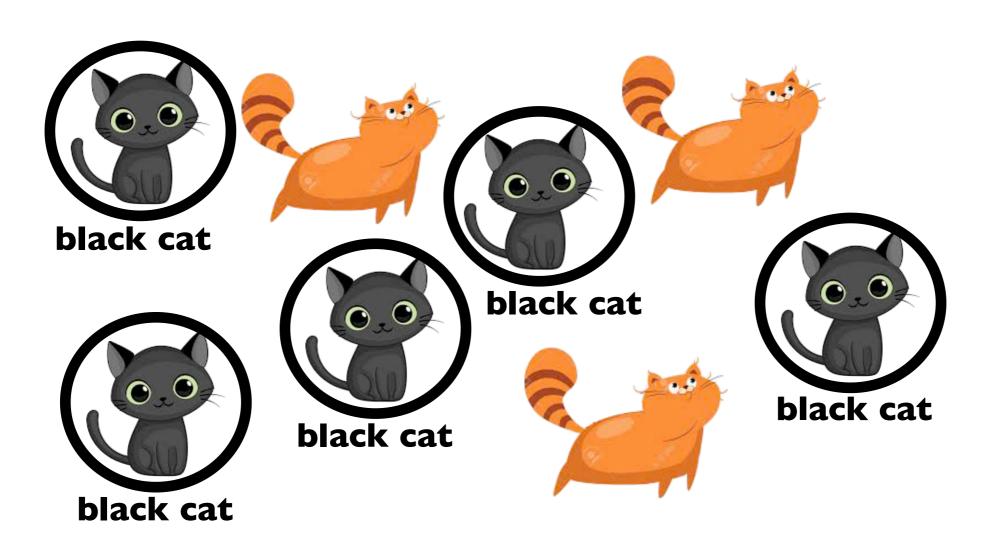
TRUEpositives

FALSEnegatives

Minimum recall = 0:

a model that does not classify any nonVERB token with category nonVERB; all classifications provided by the model do not correspond to the gold; the model provides only FALSE NEGATIVES and no TRUE POSITIVES.

Recap: precision and recall



The model has to classify the category **black cats** in the set above. It is important that it classifies correctly each cat as black (precision), but also that it classifies all the cats that are black (recall)! All and only the black cats!

F=Precision+Recall

Putting together **precision** and **recall** we have a better evaluation in which it is shown how many words are correctly classified: it is called **FI-score** and is calculated as the ARMONIC MEAN of Precision and Recall:

$$precision x recall$$
 $FI = 2 x$
 $precision + recall$