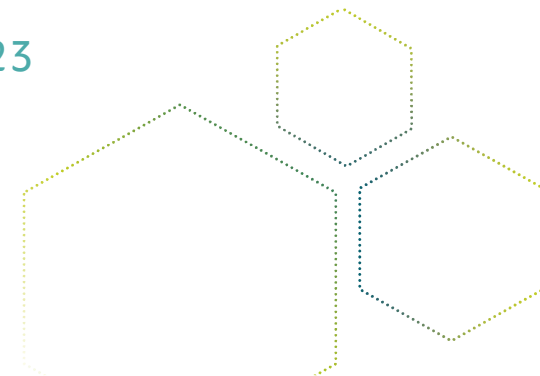

IA-based tools for patent classification

IPC Committee of Experts 54 – February, 22 and 23, 2023

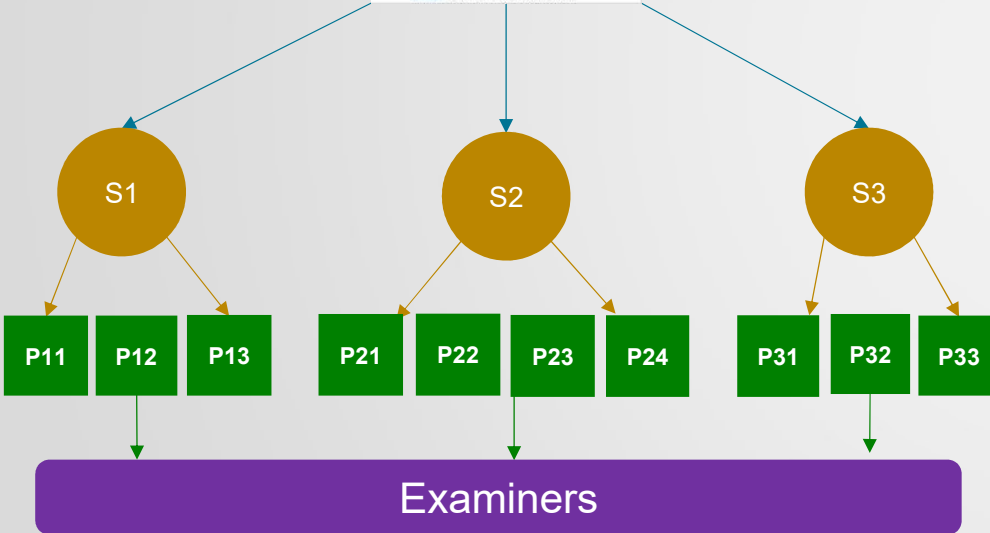


DISPATCHING OF PATENT APPLICATIONS

A TIME CONSUMING TASK

ca. 300 applications per week

Servicio	Reservat	Titre del subvenció
1	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
2	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
3	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
4	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
5	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
6	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
7	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
8	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
9	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
10	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
11	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
12	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
13	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
14	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
15	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
16	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
17	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
18	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
19	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
20	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
21	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
22	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
23	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
24	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
25	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
26	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
27	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
28	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
29	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
30	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
31	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
32	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
33	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
34	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
35	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
36	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
37	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
38	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
39	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
40	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
41	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
42	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
43	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
44	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
45	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
46	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
47	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
48	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
49	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
50	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
51	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
52	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
53	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
54	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
55	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
56	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
57	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
58	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
59	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
60	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
61	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
62	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
63	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
64	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
65	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
66	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
67	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
68	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
69	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
70	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
71	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
72	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
73	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
74	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
75	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
76	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
77	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
78	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
79	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
80	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
81	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
82	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
83	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
84	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
85	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
86	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
87	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
88	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
89	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
90	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
91	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
92	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
93	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
94	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
95	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
96	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
97	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
98	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
99	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR
100	11111111	AVANCE DE TRABAJO PARA EL SERVIDOR



Before 2019

A two-step manual process :

- ✓ **Pre-distribution:**
By Service Managers
- ✓ **Distribution to examiners:**
By Team Managers



PATENT PRECLASSIFICATION

A BASIC FIRST MODEL

- ▶ **Data collection** with 150 000 patents published by INPI over a 10 years period
- ▶ **Model development** using automatic language processing and learning
- ▶ **In production** in April 2019 replacing manual dispatching
- ▶ Provides the most likely examiners' team

In production model accuracy : 83%





PATENT PRECLASSIFICATION

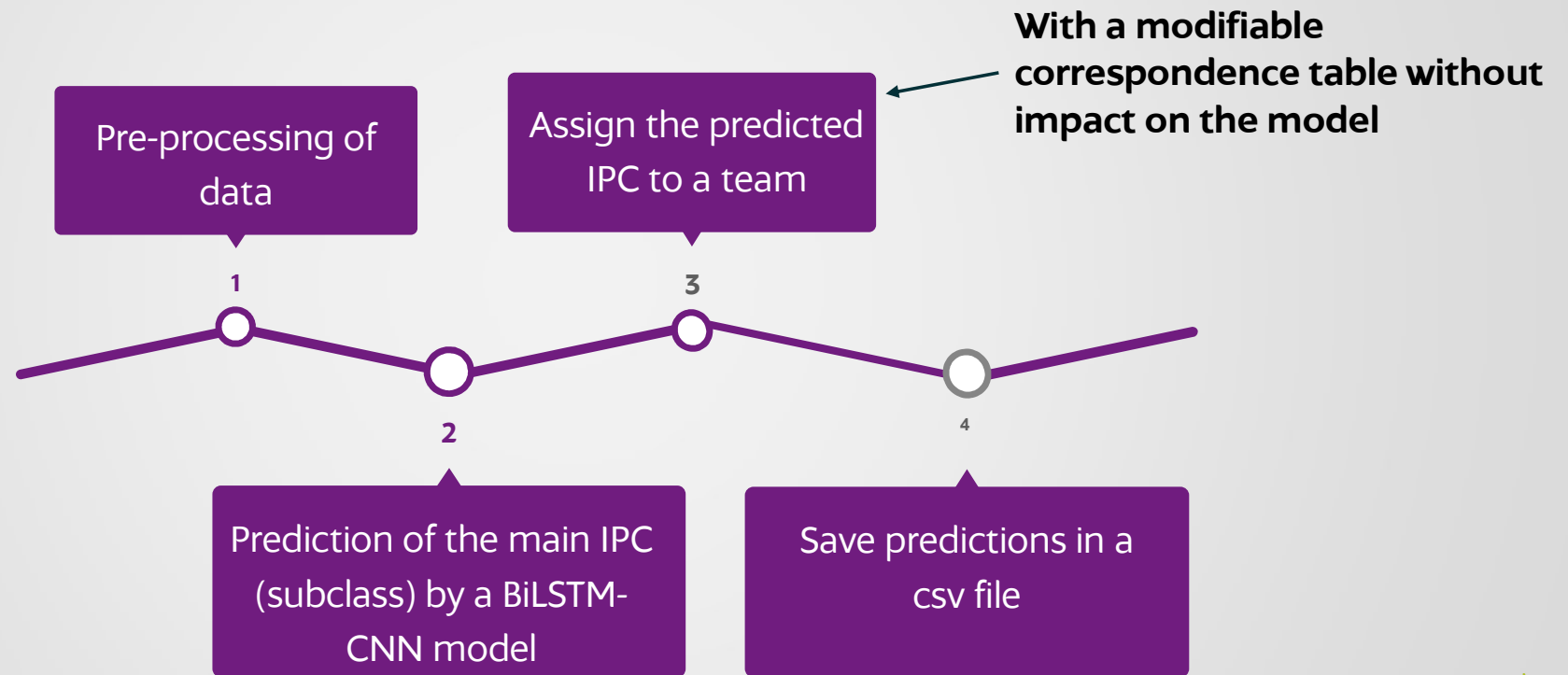
ADAPTATION OF THE FIRST MODEL

- ▶ **Reorganisation of the Patent Department mid 2019** with the transition to 10 examiners' teams
- ▶ **Need to re-train the model** by targeting these new teams
- ▶ **Impact on performances** with a significant decrease of accuracy

In production model accuracy : 74,5%

PATENT PRECLASSIFICATION

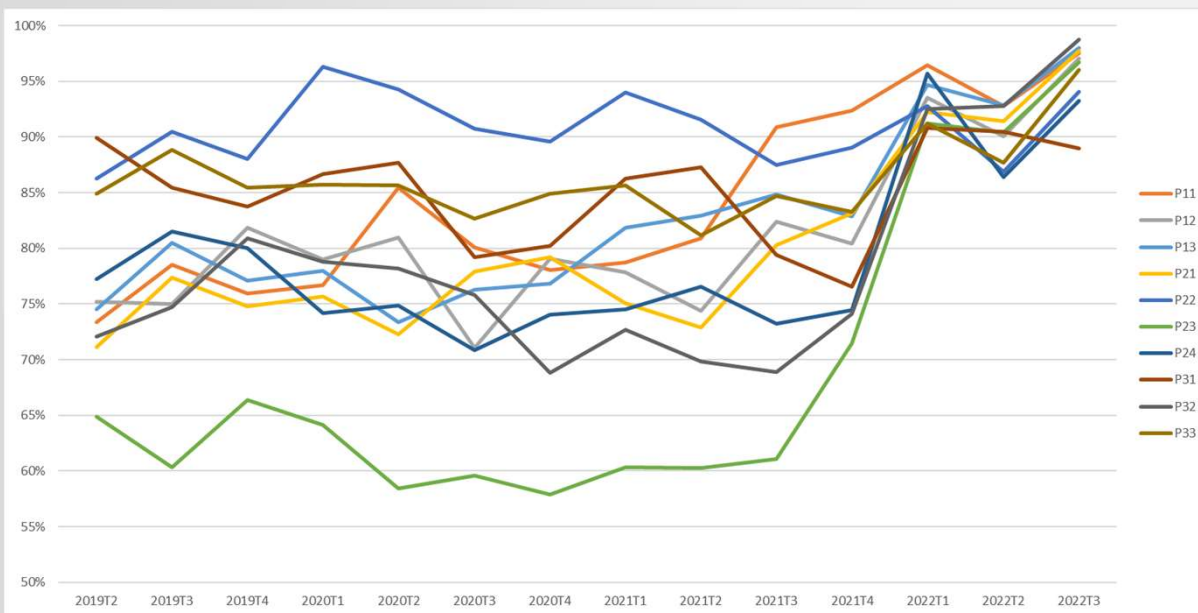
A NEW PROCESS ROBUST TO CHANGES



PATENT PRECLASSIFICATION

SIGNIFICANT INCREASE IN PERFORMANCE

- ▶ **Adding recent data** to the original data for training and evaluation
- ▶ **Production start** in December 2021 with IPC (subclass) and predicted pole



Team	Average correct prediction rate old model	Average correct prediction rate new model
P11	81%	96%
P12	78%	94%
P13	79%	95%
P21	76%	94%
P22	91%	91%
P23	62%	93%
P24	76%	92%
P31	84%	90%
P32	74%	95%
P33	85%	92%

Source: DB



PATENT CLASSIFICATION

A COMPLEX TASK

- ▶ Each patent application can have **several classifications** : multi-label classification (**XML : Extreme Multi-Label Learning**)
- ▶ **Extrem classification** requires a sustained effort and sophisticated models
- ▶ Preliminary feasibility study by **evaluating existing approaches** to classification at a subclass level and at a subgroup level(**XML-CNN, Paralabel, BERT**)
- ▶ **Collaboration** of the Lab INPI with a research institute (INRIA)

PATENT CLASSIFICATION

DATA PREPARATION

- ▶ Need for good quality and quantity of **data**
- ▶ **Collection** of a large volume, **increase** and **structuring** in sections

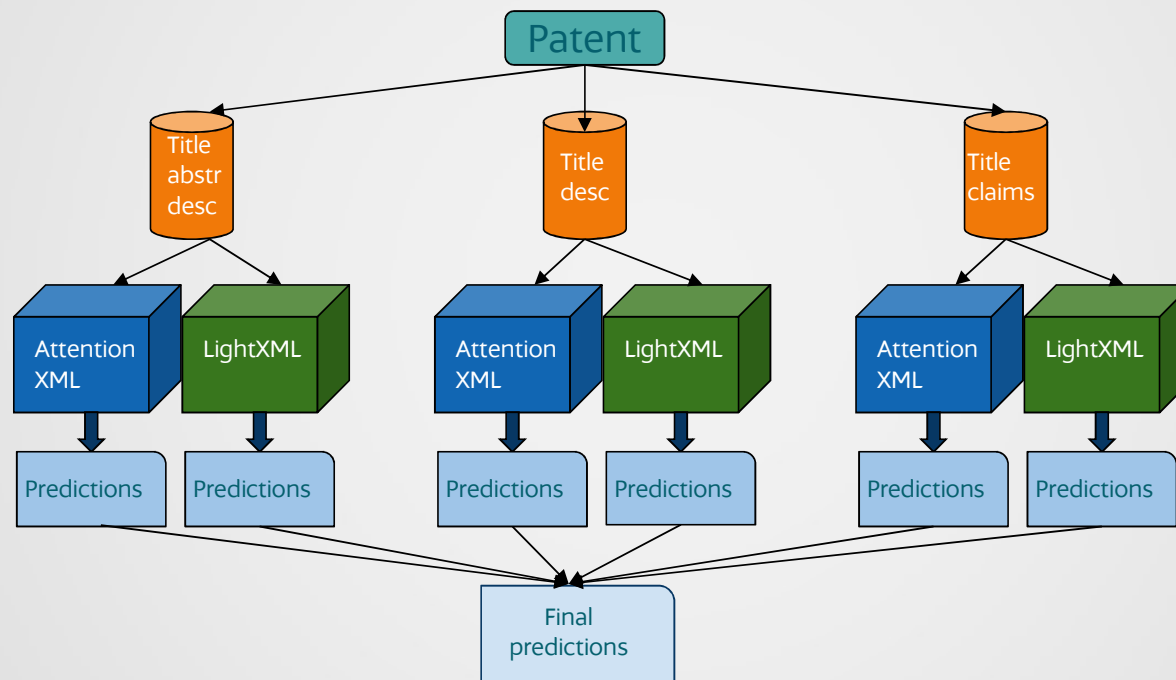
INPI data (2002-2021)

section	title	abstract	description	claims
# patents	296 270	295 421	296 216	291 539
# tokens	11	111	4202	725



PATENT CLASSIFICATION

COMBINATION OF MODELS FOR IPC (MAIN GROUP LEVEL)



Correct prediction rate : 70 %

On the first 3 predictions of the model, 1.3 rankings are correct on average



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

Thank you



www.inpi.fr



+33 (0)1 56 65 89 98



INPI France

mmathon@inpi.fr

inpi