

# IPC Union 53rd Session

INPI Brazil – Classification IT systems

***Rafael Nunes***

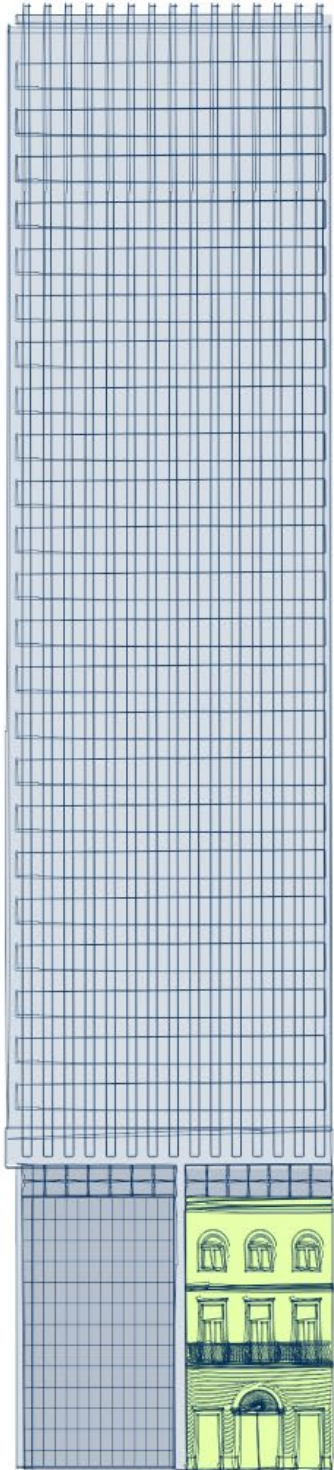
*Analyst*

*DIAPE / DIREX / PR / INPI*

Rio de Janeiro, February 24<sup>th</sup> – 25<sup>th</sup>, 2022

# *Summary*

- IPC-related IT systems
- Computer-assisted classification
- AI-based IPC classification



# *IPC-related IT systems*

---

- **SINPI (INPI internal system)**
  - We have a system that support IPC annual update, it generates database scripts to update our internal system;
- **IPC classification**
  - We don't have any automated system for IPC classification
- **IPC reclassification**
  - We use a Python script to evaluate IPC reclassification needs and generate corresponding database update script using Results XML as source information

# *Classification-related IT systems*

---

- **IPC PUB**
  - Portuguese version of IPC since 2018
  - Helps specialists on classifying patent applications
- **CPC Scheme Updater**
  - Automatically monitor new versions and update our systems to make new CPC available to use for our personnel

# *AI-based classification*

---

- Before 2018 | E-Distribuidor (NN)
  - About 30% mean accuracy @ 2018
  - Proprietary technology (matlab) compiled to executable on a non-repeatable process
  - Retrain not possible by INPI alone

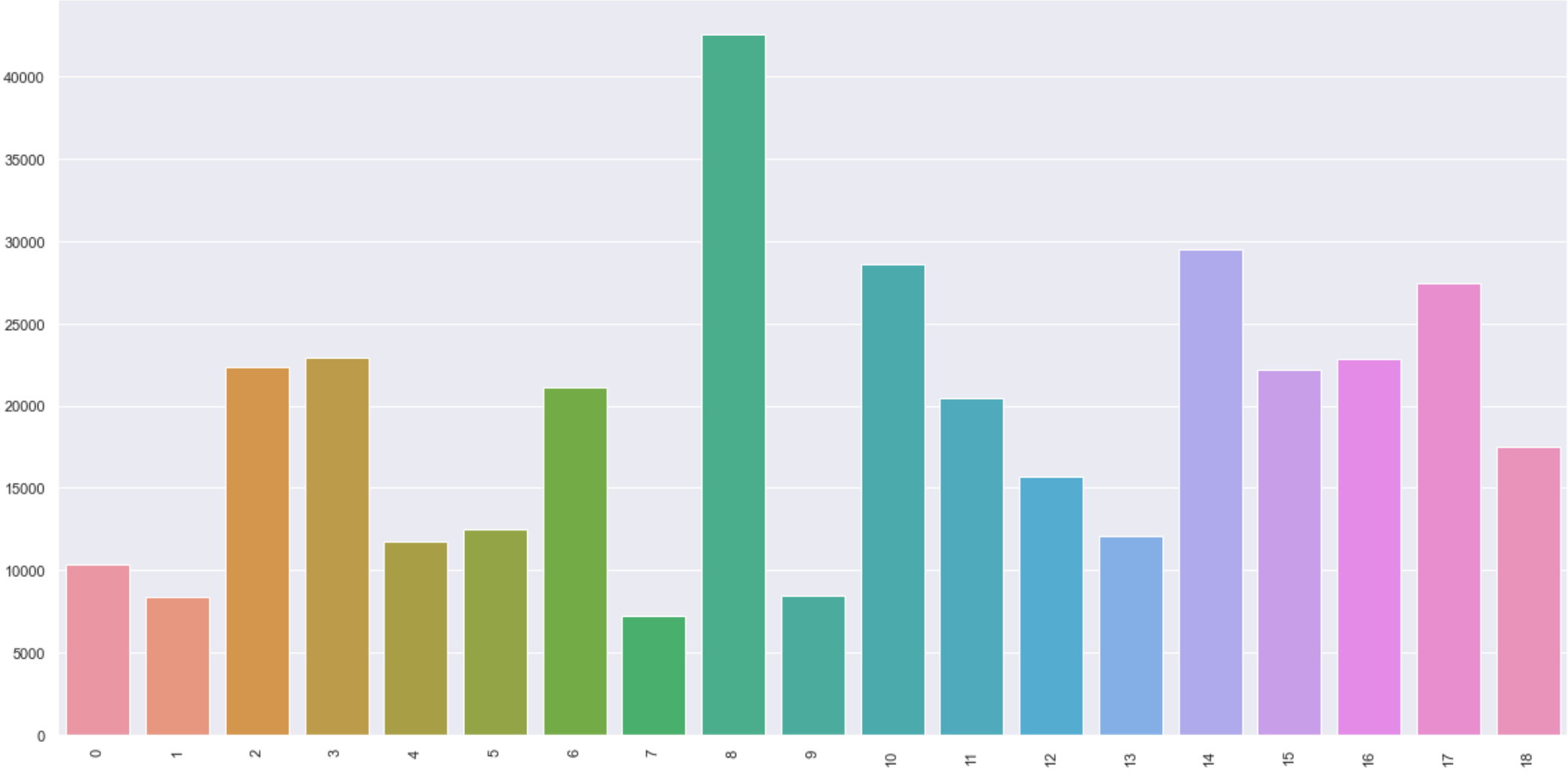
# *AI-based classification*

---

- After March 2018 | Application Router
  - Based on Python open source implementation of scikit-learn (sklearn)
  - Using Title and Abstract text combined
  - Using TF-IDF and LSA (1,000 dims) to extract features
  - Trained a Multilayer Perceptron NN on a +360k examples dataset
  - To suggest 1 out of 19 exam area

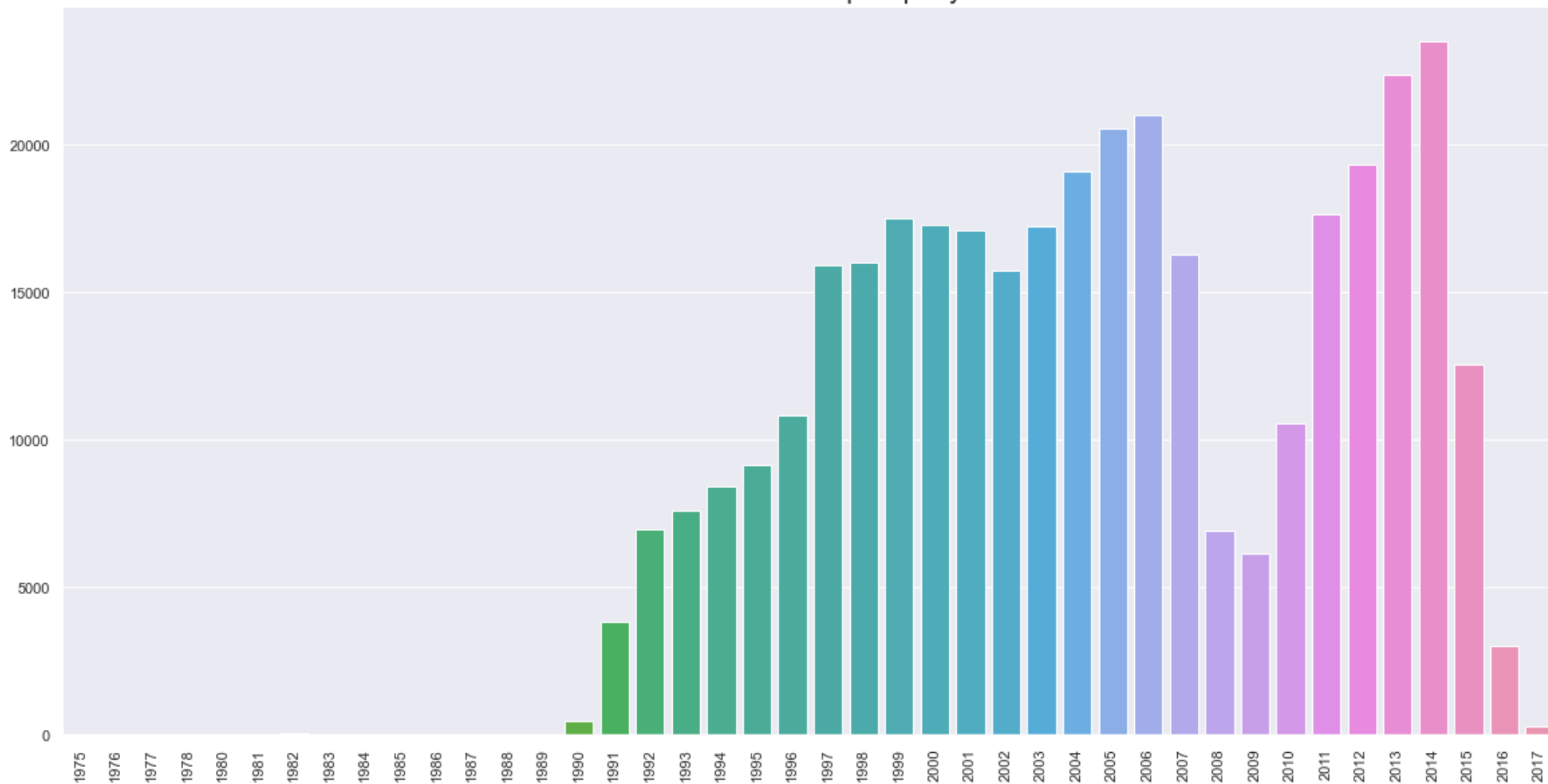
# AI-based classification

Train dataset examples per class



# AI-based classification

Train dataset examples per year





# *AI-based classification*

---

- Application Router – benefits
  - From April/2018 to June/2019 Application Router routed correctly about 3.000 patent applications more than e-Distribuidor (former solution), it **performed 90% better**. e-Distribuidor would routed correctly 3,342 applications and Application Router routed correctly 6,381 application
  - Mean Accuracy 56,85% @ 2019
  - Repeatable training process

# *AI-based classification*

---

- Future developments
  - Update classifier training
  - Research new algorithms
  - Investigate new features techniques

# Thank you!



*[rafael.nunes@inpi.gov.br](mailto:rafael.nunes@inpi.gov.br)*

*[www.inpi.gov.br](http://www.inpi.gov.br)*