# Juan Cruz Sosa

## Personal Information


| | | | |
|---|---|---|---|
| Phone: | **1166648875** | Email: | juan.cruz.sosa.92@gmail.com |
| Git repository: | github.com/nirvguy | Kaggle User: | kaggle.com/jcsosa |
| LinkedIn: | juan-cruz-sosa | | |


## Education

From 2012 - Present **University - Licenciatura en Ciencias de la Computación (Bachelor of Computer Science)**, *Facultad de Ciencias Exactas y Naturales (UBA)*, GPA: 8.21/10. Expected to graduate in Dec. 2019.

## Work Experience

From 2018–Present **Machine Learning Engineer Ssr**, Workia Solutions.

Development of several Python Natural Language Processing pipelines at Workia Solutions to enhance HR workflows in two of their major projects.

At Birpin project (a Job Search platform powered by IA) I

- Improve existing models for Resume Parsing & Analysis using the Spacy framework pipeline for Spanish.
- Curriculum Vitae Slicer: Design, Development and Training from scratch a Neural Network for predicting change of contexts in a curriculum vitae (such as change of work experience or study blocks) using multiple text feature engineering techniques on adjacents text lines of resume text.
- Job title text compression using custom fuzzy techniques.
- Training a domain-specific Word2Vec and Doc2Vec spanish language models of text gathered in jobs description.
- Design of Model for Job hierarchy and career segmentation.
- Improve automated offer matching to candidates based in their profile content and parsed resume using previously trained Text Embeddings for every job experience.
- Collaborative filtering: Implementation of Cypher queries to perform matching based in graph neighbour analysis of offers and users behaviour.

At HCWork (a HR for employee managment) I

- Build a chatbot powered by DialogFlow for the employees self-service module
- Develop a module for Anomaly Detection using Metabase reporting.

### Freelance

From 2014–Present **Web development in Django**, Instituto San Juan Bosco, colegiosanjuanbosco.edu.ar.

## Github Projects

Python - PyTorch **Torchtrainer**, *Framework for training/validation of models written in PyTorch*.
Development of several extensible modules to facilitate training of Neural Networks and cross validation. Also provide examples of usage, documentation, and tests (with Test Driven Development).

Python - PyTorch **Annpy**, *Implementation of several neural networks types (MLP, Hopfield, RBM, DBN, GAN's)*.

Python - PyTorch **nlp**, *Projects using LSTM for Natural Language Processing tasks*.

Python - PyTorch **Kaggle-notebooks**, *Jupyter Notebooks elaborated for Kaggle competitions*.

Python **Ml-algo**, *Machine learning algorithms implementations (Linear regression, K-means clustering)*.

## Kaggle competitions

Aug. 2019 - Present **Dataton - Banco Galicia - 2019**, *URL: kaggle.com/c/banco-galicia-dataton-2019*, 106th of 191.
The task is to predict user conversion in the first quarter of 2019 using only anonymized user tracking history of 2018. With my team we rethink the problem as a time series and develop and algorithm that take different aggregated user features up to one given quarter of the 2018 and classify conversion of the next quarter. With a XGBoost model our best submission scored 83.7 % in the roc curve.

Aug. 2019 - Present **Steel Defect Detection**, *URL: kaggle.com/c/severstal-steel-defect-detection*, 650th of 906.
Image segmentation to localize and classifying surface defects of steel sheets. I divided the problem in two different tasks. The first task consists of training a binary classifier model to separate defected fixed size blocks from the non affected using Microsoft pretrained 34-layered resnet. In the another task the goal is to train the final image segmentation algorithms using U-Net convolutional architecture. Both models was develop using PyTorch.

Nov. 2017 - May 2018 **Plant Seedling Classification**, *URL: kaggle.com/c/plant-seedlings-classification*, 242th of 836.
Training of models to perform classification of plant species given a photo. Experimented with multiple types of Convolutional Networks (from simple 2-conv layers to a deep resnet network) with PyTorch and Keras too. The best submission scored 242th of 836 with 97.35 % classification accuracy using Data augmentation and Transfer Learning on a resnet of 34 layers.

## Technical Expertise & Programming Skills

| | |
|---|---|
| Languages | Python (5 years) \| C/C++ (5 years) \| SQL (6 years) \| Matlab (at University) |
| Databases | PostgreSQL \| MongoDB \| Neo4j (used for collaborative filtering) |
| VCS | Git (4 years) |
| NLP | Spacy \| nltk \| gensim |
| Machine Learning | Scikit-learn \| PyTorch \| Keras \| Numpy \| Pandas |
| Neural Networks Architectures | MultiLayer Perceptron \| Doc2Vec \| Word2Vec \| Convolutional Networks \| LSTM \| U-Net \| Generative Adversarial Networks |

## Courses

| | |
|---|---|
| Aug. 2017 - Dec. 2017 | Deep Learning, FCEyN - 160 hours |
| March 2017- July 2017 | Introduction to Neural Networks, FCEyN - 160 hours |
| March 2016 - July 2016 | Advanced calculus, Department of Mathematics - 300 hours |
| March 2014 - July 2014 | Linear Algebra advanced course, Department of Mathematics - 300 hours |

## Languages

| | |
|---|---|
| Spanish | **Native** |
| English | **Reading & Writing - Fluent \| Speaking - Intermediate** |