# Kevin Bello Medina

**Place and Date of Birth:** Lima | 19 April 1993
**Address:** 610 Curazao Street, La Molina, Lima, Peru
**email:** ksbm190493@gmail.com

## Research Interests

- General Purpose Machine Learning: Graphical Models, Neural Networks, Deep Learning, Kernel Methods.
- Domain-Specific Machine Learning:
  - Natural Language Processing: Topic Modeling, Information Extraction, Information Retrieval, Sentiment Analysis.
  - Computer Vision: Convolutional Neural Networks for image recognition, Dimensionality Reduction.

## Career Goal

To become a fully-fledged research scientist with a multidisciplinary training. One capable of working with researchers from different disciplines and adequately integrate their skills to produce cutting-edge research for the benefit of society.

## Education

| | |
|---|---|
| Class of 2014-2 | Bachelor of Science in Mechatronics Engineering<br>**Universidad Nacional de Ingenieria**, Lima<br>*Summa Cum Laude* | Concentration: Computer Science |

## Research Experience

### Publications & Presentations

- **Bello, K.**, Cardenas, R., Coronado, A., *et al*, "Peruvian labor market demand analysis for Mechanical and Electrical Engineering", CONIMERA 21st Peruvian Congress, June 2015.
- **Bello, K.**, Cardenas, R., "Labor Market Demand Analysis for Engineering Majors in Peru Using Shallow Parsing and Topic Modeling", presented at the Machine Learning Summer School, Kyoto University, August 2015.
- Valle, A., **Bello, K.**, Cardenas, R., Villota, E., *et al*, "Analysis of the Peruvian labor market demand in the area of mining maintenance", MAPLEMIN 2015.
- Cardenas, R., **Bello, K.**, Coronado, A., *et al*, "Panorama of the Market Demand for Mechanical Engineers in South American Countries", to appear in the proceedings of the ASME International Mechanical Engineering Congress & Exposition, November 2015.

### Projects

| | |
|---|---|
| Mar. - May 2015 | **Reconceptualizing the classification of specialization programs using Topic Modeling**<br>*Undergraduate Research Assistant* | Advisor: Prof. Alberto Coronado | Detailed Information |
| Sept. - Dec. 2014 | **Recognition and classification of text chunks from job advertisements using Shallow Parsing**<br>*Undergraduate Research Assistant* | Advisor: Prof. Alberto Coronado | Detailed Information |
| May - July 2014 | **Analysis of job advertisements published on South American webpages using visualization techniques**<br>*Undergraduate Research Assistant* | Advisor: Prof. Alberto Coronado | Detailed Information |
| Oct. - Dec. 2013 | **Information retrieval system using unsupervised learning**<br>*Course Project* | Advisor: Prof. Ricardo Rodriguez | Detailed Information |
| May - Sept. 2013 | **Artificial Intelligence and Control Systems Laboratory**<br>*Course Project* | Advisor: Prof. Guillermo Kemper & Prof. Ivan Calle<br>Project 1: Localization and path-planning of a synchro and differential-drive robot<br>Project 2: Fingertip detection using edge detection and connected-component labeling<br>Project 3: Design, implementation, and control of a 2-DOF robot for real time face tracking |

## Honors and Awards

- Kyoto University's Fellowship to participate in the *Machine Learning Summer School* in Kyoto, Japan — 2015
- Highest accumulated GPA among all the students in my class($\sim$100) and top 1 percent in the last 10 years among all students at Universidad Nacional de Ingenieria ($\sim$20,000)
- Dean's list — 2010, 2011, 2012, 2013, 2014
- Honorable Mention (top 10), ACM International Collegiate Programming Contest (ICPC), South American Regional Contest by IBM — 2012, 2014

- Peruvian Council of Science and Technology Grant to participate in a summer course for computer science olympiads in Sao Paulo, Brasil                                                                                                     2013
- Ranked 35 out of 7500 participants in the Worldwide IEEExtreme Programming Competition 7.0                              2013
- *"Presidente Manuel Pardo y Lavalle Prize".* It is the highest honor given to undergraduates at Universidad Nacional de Ingenieria                                                                                                                              2012
- University of Chile's Fellowship to participate in the *1st Latin American Theoretical Informatics School*              2012
- 2nd Place in the national programming competition IEEExtreme - INTERCON, Peru                                           2012
- Certificate of recognition for achieving the highest GPA among all the students of the Mechanical Engineering Department (four programs, ~800 students) at Universidad Nacional de Ingenieria                                             2010

## COMPUTER SKILLS

| | |
|---|---|
| Programming Languages/Frameworks | Advanced: C/C++, Python, MATLAB<br>Intermediate: JAVA, R, Apache Spark, SQL, HTML, LaTeX |
| Operating Systems | LINUX [ArchLinux, Ubuntu], Windows |
| Engineering Software | Simulink, LabView, SolidWorks, AutoCAD |

## OTHER ACADEMIC ACTIVITIES

**Researcher at Artificial Intelligence and Control Systems Laboratory (GISCIA)**                    2013 - Present
Current president and member of the Artificial Intelligence and Control Systems Laboratory at Universidad Nacional de Ingenieria.

**Machine Learning Summer School**                                                                   Aug. - Sept. 2015
Kyoto University's Fellowship to participate in the MLSS'15 in Kyoto, Japan. Some of the topics included: Convex Optimization, Scalable Machine Learning, Reinforcement Learning, Large Scale Deep Learning, etc.

**MLSS 2015 Predictive Modeling Challenge**                                                           Jul. - Aug. 2015
Top 10 in the MLSS'15 predictive modeling challenge in both tasks the classification and the regression challenge.

**Course Instructor for Freshman Students at Universidad Nacional de Ingenieria**

- Introduction to Probabilistic Robotics                                                             May 2015 - July 2015
- Introduction to Machine Learning                                                                   August 2014 - December 2014
- Introduction to Algorithms                                                                         August 2013 - July 2014

**Active Contestant in Programming Competitions**
Solving math and algorithmic problems is one of my hobbies, therefore I continuously participate in online programming competitions such as Google Code Jam, TopCoder, Facebook Hacker Cup, and Codeforces.

**Summer Course for Computer Science Olympiads, Universidade Estadual de Campinas (UNICAMP)**         January 2013
Attended a three-week course about algorithms at UNICAMP, Brasil. High-quality professors from Europe and South America gave lectures to the best university teams from South America.

**1st Latin American Theoretical Informatics School (LATIN)**                                         April 2012
University of Chile's Fellowship to participate in the 1st LATIN school in which were given three lectures by professors from the Massachusetts Institute of Technology (USA), Universitat Politecnica de Catalunya (Spain), and Universidad Nacional Autonoma de Mexico (Mexico). I also attended some Plenary Lectures and Special Sessions of the LATIN Symposium.

## LANGUAGES

- English: Advanced
- Spanish: Native

## RELEVANT COURSES

**Universidad Nacional de Ingenieria**

Artificial Intelligence, Statistics and Probability, Linear Algebra, Multivariable Calculus, Digital Image Processing, Numerical Methods, Digital and Electronic Circuits, Multi-body Dynamics, Data Communication and Networks, Physics, Biology for Engineers, Differential Equations, Complex Variable.

**MOOC**  Web based courses (e.g. Coursera, EdX), scored within 90% and 100%

Neural Networks for Machine Learning (G. Hinton), Machine Learning (A. Ng), Probabilistic Graphical Models (D. Koller), Learning from Data (Y. Abu-Mostafa), Natural Language Processing (C. Manning & D. Jurafsky, M. Collins), Statistical Inference (B. Caffo), Algorithms I & II (T. Roughgarden), Programming a Robotic Car (S. Thrun)

# Research Summary

### Reconceptualizing the classification of specialization programs using Topic Modeling

**Goal:** Reconceptualize the classification of specialization programs (e.g. diploma) using unstructured text from job advertisements.
**Significance:** Reveal latent connections among the majors with highest demand in the Peruvian labor market. Aiming to propose a new integration of specialization programs offered by Peruvian universities so that future professionals can be more competitive.
**Methods:** Six mixed-membership models were compared in order to observe the effect over model dimensionality, all of the models used Latent Dirichlet Allocation (LDA). The data source and the learning algorithm varied for each model. For the case of data source, two variations, one using the whole text, and the other using text only from functions and requirements of the job advertisement (a classifier of a previous work was used. See the description of the project below). And for the case of the learning algorithms, three variations. Two using Variational Expectation-Maximization, one with the hyperparameter being estimated and the other fixed, and the remaining with Gibbs Sampling. The 10-fold cross-validation method was used for model selection in order to find the optimal number of latent categories. The six mixed-membership models were implemented using R and C++.
**Results:** The analysis revealed several major clusters, which means that majors in the same cluster are required almost indistinctly, and that specializations converge to a mixture of common pools of knowledge and required skills. This suggests that some specialization programs should be a mixture of more than two different fields. Currently, specialization programs in Peru combine at most two different fields.

### Recognition and classification of text chunks from job advertisements using Shallow Parsing

**Goal:** Recognize and classify text chunks such as functions, requirements, majors, and so forth from job advertisements published in Peruvian websites.
**Significance:** After developing the classifier and running over hundreds of thousands job advertisements, the results will shed light the requirements and functions with highest demand for each major. Thus knowing these statistics will make an impact in the study programs taught by educational institutions in Peru.
**Methods:** The pre-processing step consisted of Natural Language Processing techniques such as regular expressions, word tokenization, and sentence segmentation. Then the problem was reduced to a sequence labeling problem (Shallow Parsing) in which the base phrases defined were: *Major* (major required for the position, Noun Phrase), *Requirements* (list of skills the applicant needs to have, Noun Phrase and Verbal Phrase), and *Functions* (list of tasks the applicant will have to accomplish, Verbal Phrase). A generative model using Hidden Markov Models and a discriminative model using Averaged Structured Perceptron were compared.
**Results:** The results were close to the state of the art performances in the Spanish language, with a 87.1% F1 score for the *Major* predictions.

### Analysis of job advertisements published on South American webpages using visualization techniques

**Goal:** Perform a descriptive analysis of online South America job advertisements using visualization techniques.
**Significance:** Providing statistics about the majors with highest demand can help people to get a notion of the current state of industry. Furthermore, people could observe and decide whether some career paths are viable to pursue.
**Methods:** A web crawler using the Scrapy Python framework was implemented. In addition, the job advertisements were stored in a SQL-model database. A detector (to determine the majors required in the advertisement) was developed using regular expressions. The eighty most popular majors were considered in the detection phase. The statistics were shown through an interactive visualization by using the Javascript D3 library.
**Results:** The detector system achieved a precision of 96% in average. Analysis of the results are explained in the website www.empleatron.com (Spanish). A Treemap showing the distribution of the labor demand, and a Circleplot showing the relationship among the majors are presented in the webpage.

### Information retrieval system using unsupervised learning

**Goal:** Provide an extra feature to a classical information retrieval (IR) system using clustering.
**Significance:** The possibility of retrieving documents that are not relevant for a classical IR system. These additional retrieved documents are considered similar, and so potential relevant results, because they belong to the same cluster that the documents provided by a classical IR system.
**Methods:** First, documents were modeled as high-dimensional vectors (bag of words) with the TF-IDF weighting method used over the components. Then the Singular Value Decomposition algorithm was applied for dimensionality reduction, and the similarity of documents was defined by the cosine similarity. Finally, the Spherical K-means algorithm was implemented in the Python language to perform clustering.
**Results:** The outcomes showed that some documents can be relevant for a given search query even without containing any of the words from the search query.