Denzil Correa, Ph.D.

correa.denzil@gmail•com
https://correa.in/
http://www.linkedin.com/in/denzilcorrea/

April 2019

Skills

Leadership	Scientific Guidance, Technology Strategy, Stakeholder Communication
Business	Business Case Development, Customer Understanding, Market Positioning
Research	Natural Language Processing, Machine Learning, Text Mining, Network Science
Research Areas	Social Computing, Computational Biology, Mining Software Repositories

Education

2009 – 2014 **Ph.D, IIIT-Delhi, India** in Computer Science
Thesis title: *Content Quality in Web 2.0 Services - Analysis, Detection, Systems and Enhancement*

2012 **M.Tech., IIIT-Delhi, India** in Computer Science
Thesis title: *WhACKY! What Anyone Could Know About You from Twitter.*

2003 – 2007 **B.Engg. (Hons), University of Mumbai, India** in Computer Science
First Class Honours with Distinction.
Thesis title: *Musical Information Retrieval Systems for Instrumental Audio.*

Awards and Achievements

Awards

2016 **Young Researcher Scholarship**, Heidelberg Laureate Forum, Germany
2015 **Post Doctoral Research Fellowship**, Max Planck Gesellschaft, Germany
2011 **Yahoo! Summer School Fellowship**, Yahoo!, India
2010 **Research Fellowship**, National Internet Exchange (NIXI), India
2009 **Research Fellowship**, Government of India

Patents

2013 **System and Method for Verifying Credentials** US20130275753A1

Employment History - Roles and Responsibilities

2018 ... **Senior Data Scientist** for Bayer Pharma, Germany
- Interpret Business Needs and Formulate Key Business Metrics for Pharma R&D
- Initiate and Prioritize Data Science Project Business Cases to Generate Asset Value and Decrease Time-to-Market
- Lead Data Science Projects and Mentor Junior Data Scientists to Develop Novel Scientific Approaches
- Interface between Business and Technology Teams to Ensure Shared Common Objectives
- Communicate Key Insights and Prototypes to Key Business Stakeholders to enable Actionable Business Decisions
- Contribute to Data Science Oriented Strategy to Achieve Organizational Business Goals
- Champion for Data Science Benefits to Organizational Business Needs

2016 – 2017 **Data Scientist** for Bayer Pharma, Germany
- Develop Novel Data Science Algorithms for Pharma R&D to Decrease Time-to-Market
- Communicate Key Insights and Prototypes to Key Business Stakeholders to Make Actionable Decisions

2015 – 2016 **Post Doctoral Research Fellow** at Max Planck Institute for Software Systems, Germany
- Develop Novel Data Science Algorithms and Techniques to Study Cultural Attitudes and Perceptions
- Generate Unique Insights for User Behavior on Social Media
- Scientific Mentor to Ph.D students and Research Interns

2014 **Research Intern** at Max Planck Institute for Software Systems, Germany
- Develop Novel Computational Methods to Discover Key Insights for User Behavior from Anonymous Social Media

2012 **Research Intern** at School of Information Systems, Singapore Management University, Singapore
- Develop Novel Computational Methods to Increase User Engagement on Mobile Social Networks

2010 **Research Intern** at IBM Research, India
- Developed a Novel Machine Learning Algorithm to Detect Spam on Social Media with 80% accuracy

2007 – 2009 **Research Engineer** at Infosys Research, India
- Implement Novel Research Machine Learning, Data Mining and Natural Language Processing Algorithms to generate Business Value for Industries like Retail, Finance and Aeronautics

Industry Projects

2019 ... **Deep Learning for Automated Generation of Regulatory Documents**
Business Value 42M€ per year
Developing a Novel Recurrent Neural Network (RNN) to generate first drafts of Regulatory Documents like Clinical Study Reports
Research Natural Language Processing, Machine Learning
Leadership Customer Understanding, Business Case, Scientific Guidance, Stakeholder Communication

2019 **Deep Learning for Malaria Prediction from Cell Images**
Developed a Novel Convolution Neural Network (CNN) based on transfer learning to detect malaria from cell images with 99% accuracy
Research Machine Learning

2018 ... **Deep Learning for Prediction of Combinatorial Drug Efficacy**
Potential Business Value 0.5 BN€ per year
Developing a Novel Graph Neural Network to predict Drug-Drug efficacy. Our work aims to provide direct value for concomitant therapies in various areas.
Research Network Science, Machine Learning
Leadership Market Positioning, Business Case, Scientific Guidance, Stakeholder Communication

Deep Learning for Prediction of Combinatorial Drug Side Effects
Potential Business Value 80M€ per year
Developing a Novel Graph Neural Network to predict side effects of drugs when taken in combination but not observed in solitary usage
Research Network Science, Machine Learning
Leadership Market Positioning, Business Case, Scientific Guidance, Stakeholder Communication

Prediction of Drug Safety from Pre-Clinical Research and Biological Data
Potential Business Value 100M€ per year
Developing a Novel Ensemble Machine Learning algorithm based on Random Forests that utilizes chemical properties from pre-clinical research and biological data to predict clinical safety of compounds
Research Network Science, Machine Learning
Leadership Business Case, Scientific Guidance, Stakeholder Communication

2017 **Identification of Patient Concerns from Social Media**
Potential Business Value 2M€
Implemented a state-of-the-art character Recurrent Neural Network (char RNN) architecture to identify patient concerns from customer data with 91% accuracy
Exploratory Network Analysis to find Key Opinion Leaders and Information Flow with 97% accuracy
Research Natural Language Processing, Machine Learning, Social Computing
Leadership Scientific Guidance

Industry Projects (continued)

Deep Learning for Language Translation of Regulatory Labels Documents
Business Value 0.5M€ per year
We used Recurrent Neural Networks (RNNs) to translate drug package label documents between 5 European languages with BLEU score of 84%
Research Natural Language Processing, Machine Learning

2010 **Automated Spam Detection on Social Media**
Implemented a Bayesian Machine Learning framework to detect spam on social media with 80% accuracy
Research Natural Language Processing, Machine Learning

2008 **Tabular Extraction from Financial Documents**
Business Value 0.5M€ per year
Implemented an SVM based machine learning algorithm to detect and extract tabular (structured) information from unstructured data with 77% accuracy
Software Engineering Java, Testing

Academic Projects and Research Publications

Journal Articles

1. Mondal, M., Silva, L. A., Correa, D., & Benevenuto, F. (2018). Characterizing usage of explicit hate expressions in social media. *New Review of Hypermedia and Multimedia*, *24*(2), 110–130.
2. Agarwal, S., Mittal, N., Katyal, R., Sureka, A., & Correa, D. (2016). Women in computer science research: what is the bibliography data telling us? *ACM SIGCAS Computers and Society*, *46*(1), 7–19.
3. Correa, D. & Sureka, A. (2013c). Solutions to detect and analyze online radicalization: a survey. *arXiv preprint arXiv:1301.4916*.

Conference Proceedings

1. Correa, D., Sureka, A., & Lal, S. (2017). Investigation of ir based topic models on issue tracking systems to infer software-specific semantic related term pairs. In *2017 tenth international conference on contemporary computing (ic3)* (pp. 1–5). IEEE.
2. Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016). Analyzing the targets of hate in online social media. In *Tenth international aaai conference on web and social media*.
3. Xia, X., Lo, D., Correa, D., Sureka, A., & Shihab, E. (2016). It takes two to tango: deleted stack overflow question prediction with text and meta features. In *2016 ieee 40th annual computer software and applications conference (compsac)* (Vol. 1, pp. 73–82). IEEE.
4. Correa, D., Silva, L. A., Mondal, M., Benevenuto, F., & Gummadi, K. P. (2015). The many shades of anonymity: characterizing anonymous social media content. In *Ninth international aaai conference on web and social media*.
5. Correa, D. & Sureka, A. (2014). Chaff from the wheat: characterization and modeling of deleted questions on stack overflow. In *Proceedings of the 23rd international conference on world wide web* (pp. 631–642). ACM.

6 Lal, S., Correa, D., & Sureka, A. (2014). Miqs: characterization and prediction of migrated questions on stackexchange. In *In proceedings of the 21st asia-pacific software engineering conference* (p. 9).

7 Correa, D., Lal, S., Saini, A., & Sureka, A. (2013). Samekana: a browser extension for including relevant web links in issue tracking system discussion forum. In *2013 20th asia-pacific software engineering conference (apsec)* (Vol. 1, pp. 25–33). IEEE.

8 Correa, D. & Sureka, A. (2013a). Fit or unfit: analysis and prediction of 'closed questions' on stack overflow. In *Proceedings of the first acm conference on online social networks* (pp. 201–212). ACM.

9 Correa, D. & Sureka, A. (2013b). Integrating issue tracking systems with community-based question and answering websites. In *2013 22nd australian software engineering conference* (pp. 88–96). IEEE.

10 Lim, E.-P., Correa, D., Lo, D., Finegold, M., & Zhu, F. (2013). Reviving dormant ties in an online social network experiment. In *Seventh international aaai conference on weblogs and social media*.

11 Correa, D., Sureka, A., & Pundir, M. (2012). Itop: interaction based topic centric community discovery on twitter. In *Proceedings of the 5th ph. d. workshop on information and knowledge* (pp. 51–58). ACM.

12 Correa, D., Sureka, A., & Sethi, R. (2012). Whacky!-what anyone could know about you from twitter. In *2012 tenth annual international conference on privacy, security and trust* (pp. 43–50). IEEE.

13 Correa, D. & Sureka, A. (2011). Mining tweets for tag recommendation on social media. In *Proceedings of the 3rd international workshop on search and mining user-generated contents* (pp. 69–76). ACM.

14 Sureka, A., Goyal, V., Correa, D., & Mondal, A. (2010). Generating domain-specific ontology from common-sense semantic network for target-specific sentiment analysis. In *Computer and information science, artificial intelligence paper*.

15 Sureka, A., Goyal, V., Correa, D., & Mondal, A. (2009). Polarity classification of subjective words using common-sense knowledge-base. In *International workshop on rough sets, fuzzy sets, data mining, and granular-soft computing* (pp. 486–493). Springer, Berlin, Heidelberg.