PALLIKA H KANANI
Lead Data Scientist, Oracle Labs
http://www.linkedin.com/in/pallika

413-695-1399
pallika.kanani@oracle.com

SUMMARY

- Build and lead a team of data scientists within the Machine Learning Research Group at Oracle Labs. Work closely with product groups across Oracle on applications of Machine Learning with strong business impact. Application domains include Sales, Marketing, HR, Retail, Finance, Healthcare, and Social Media Analytics.

- PhD in Computer Science from UMass, Amherst, with research interests at the intersection of Natural Language Processing and Machine Learning. Interested in information extraction, semi-supervised and transfer learning, fairness and privacy, and active information acquisition.

- Senior Advisory Board Member for Women in Machine Learning, a premier international organization to promote women in Machine Learning

WORK EXPERIENCE

- **Data Science Lead, Machine Learning Research Group, Oracle Labs (11/16-Present)**
  Lead a team of data scientists within the Machine Learning Research Group at Oracle Labs. Work closely with product groups across Oracle on applications of Machine Learning with strong business impact. Application domains include Sales, Marketing, HR, Retail, Finance, Healthcare, and Social Media Analytics. Also work on core research problems in Fairness and Privacy in Machine Learning. Responsible for building and recruiting for the team, setting the technical direction of the projects, providing research guidance to the data scientists, coordinating with cross-functional stakeholders, and reporting progress to senior executives regularly.

- **Research Scientist, Oracle Labs (03/12 - 11/16)**
  Worked on multiple core and applied ML research projects in collaboration with various product groups. Activities included business problem understanding, data collection, research, modeling, and technology transfer. Multiple research projects resulted in either new products or replacing third party solutions in deployed systems. Also worked on patenting and publishing some of this research. Supervised research interns, represented the company at external conferences, helped organized internal ML events to build a company-wide community, and made presentations to high level execs on Machine Learning to help shape ML strategy.

- **Graduate Research Assistant, UMass, Amherst (01/05 – 02/12)**
  Worked with Prof. Andrew McCallum in the Information Extraction and Synthesis Lab. My interests include data mining, information extraction, machine learning and NLP. Worked on Author Coreference for scientific paper search engine, Rexa. In many scenarios, functionality or accuracy of machine learning systems can be improved using additional information. My thesis was about efficiently querying and incorporating external information under resource constraints.

- **Research Intern, IBM T.J. Watson Research Center, New York(09/09 – 03/10)**
  Worked on the Jeopardy! Challenge as part of the IBM DeepQA project. Worked on applying formal machine learning approaches to some of the existing components of the system. The proposed method was incorporated in the final system.

- **Summer Intern, Microsoft Research, Redmond (06/09 – 08/09)**
  Interned with the Text Mining, Search and Navigation group. I worked on developing proximity features to improve retrieval performance on queries with many words or rare queries in Bing. One of the approaches that I worked on was shipped for production.

- **Summer Intern, IBM T.J. Watson Research Center, New York(05/08-08/08)**
  Designed and implemented prediction-time active information acquisition systems for customer targeting. We tested strategies for selecting a subset of instances for which to acquire additional information, so as to get most improvement in accuracy at prediction time with least cost.

- **Project Consultant to White Oak Technologies Inc. Maryland (05/07 – 10/07)**
  Provided consultation for implementing a large-scale system for name disambiguation for a government contractor.

- **Business Development Manager, Neogen Chemicals Ltd. India (10/03 – 12/04)**
  Worked in the family business in various capacities, including IT infrastructure planning and execution, HR policies and administrative activities.

- **Summer Intern, Data Mining Systems Inc. New York (05/03-09/03)**
  Modified a proprietary genetic algorithm application at Deutsche Bank by analysis, design, implementation, testing and evaluation of the required functionality. The new system can analyze time series data where the time horizon is unknown or variable.

- **Software Developer, Psychology Department, NYU (05/02-05/03)**
  Worked on a system to predict "stimulus threshold of infant motor performance" using statistical models. Designed and developed interactive software, including the required database, for special experimental data analysis.

- **Software Developer, Silgate Technologies Pvt. Ltd. India (10/00-09/01)**
  Designed, developed and tested web applications.

EDUCATION

- **University of Massachusetts Amherst**   **Amherst, MA**
  Ph.D. in Computer Science, 05/2012
  Advisor: Prof. Andrew McCallum (AAAI Fellow, President of International Machine Learning Society, 50K+ citations)
  Thesis Title: Resource-bounded Information Acquisition and Learning

- **New York University**   **New York, NY**
  Master of Science in Computer Science, 09/2003

- **University of Mumbai**   **Mumbai. India**
  Bachelor of Engineering in Computer Science, 06/2000

RESEARCH INTERESTS

- Data Science, Machine Learning, Natural Language Processing, information extraction, semi-supervised and transfer learning, model interpretability, and active information acquisition.

CITATIONS

- Number of citations (October 2019): 388
- h-index: 9
- i10-index: 9
- Three papers with more than 50 citations

PUBLICATIONS

- Peterson, D., Kanani, P., Marathe, V. "Private Federated Learning with Domain Adaptation", NeurIPS 2019 Workshop on Federated Learning for Data Privacy and Confidentiality

- Tristan, J.-B., Panda, S., Mahmoudian, H., Kanani, P., Wick, M., "Using Bayes Factors to Control for Fairness A Case Study on Learning to Rank", NeurIPS 2019 Workshop on Robust AI in Financial Services: Data, Fairness, Explainability, Trustworthiness and Privacy

- Wick, M., Kanani, P., Pocock, A. "Minimally Constrained Multilingual Word Embeddings via Artificial Code Switching", AAAI 2016, also in NIPS Workshop on Multi-Task and Transfer Learning 2016

- Kanani, P., Wick, M., Pocock, A. "Attribute Extraction from Noisy Text Using Character-based Sequence Tagging Models", NIPS Workshop on Machine Learning for e-Commerce 2016.

- Kanani, P. "Resource-bounded Information Acquisition and Learning" PhD Thesis 2012, University of Massachusetts Amherst.

- Kanani, P. and McCallum, A. "Selecting Actions for Resource-bounded Information Extraction using Reinforcement Learning" In Proceedings WSDM 2012.

- Kanani, P. and McCallum, A. "Learning to Select Actions for Resource-bounded Information Extraction". UMass Technical Report 2011, UM-CS-2011-042.

- Svore, K., Kanani, P., Khan, N., "How good is a span of terms? Exploiting proximity to improve web retrieval", Proceedings of the 33rd ACM SIGIR, 2010, pp. 154-161.

- Kanani, P., McCallum, A. and Hu, S., "Resource-bounded Information Extraction: Acquiring Missing Feature Values On Demand", Proceedings of the 14th PAKDD, 2010 (**Winner of Best Student Paper Runner Up Award**). Also appeared as poster in NESCAI 2010 and WiML, 2010

- Kanani, P., McCallum, A., and Sitaraman, R., "Towards Theoretical Bounds for Resource-bounded Information Gathering for Correlation Clustering", UMass Tech. Report 2009, UM-CS-2009-027.

- Kanani, P. and Melville, P., "Prediction-time Active Feature-value Acquisition for Customer Targeting", NIPS 2008 Workshop on Cost Sensitive Learning.

- Kanani, P. and McCallum, A., "Efficient Strategies for Improving Partitioning-Based Author Coreference by Incorporating Web Pages as Graph Nodes," Proceedings of the Workshop on Information Integration on the Web, 2007 pp. 38-43. Also appeared as a poster in NESCAI 2007.

- Culotta, A., Kanani, P., Hall, R., Wick, M. and McCallum, A., "Author Disambiguation using Error-driven Machine Learning with a Ranking Loss Function," IIWeb 2007

- Kanani, P. and McCallum, A., "Resource-bounded Information Gathering for Correlation Clustering," Proceedings of COLT 2007, Open Problems Track, LNAI 4539, pp. 625–627.

- Kanani, P., McCallum, A. and Pal, C., "Improving Author Coreference by Resource-bounded Information Gathering from the Web," Proceedings of the IJCAI 2007, pp. 429-434

- Young, J.W., Webster, T., Adolph, K. E., Robinson, S. R., & Kanani, P. "The effects of sampling frequency on developmental trajectories". International Society for Developmental Psychobiology, New Orleans, LA, 2003

- Biu, O., Young, J. W., Pethkongkathon, J., Kanani, P., & Adolph, K. E. "A microgenetic analysis of the trajectory of motor development". Society for Research in Child Development, Tampa, FL, 2003

- Young, J.W., Biu, O., Pethkongkathon, J., Kanani, P., & Adolph, K.E., "Continuity and discontinuity in motor skill acquisition". ISDP, Orlando, FL, 2002

PATENTS

- Privacy Preserving Collaborative Learning With Domain Adaptation
  PH Kanani, D Peterson, VJ Marathe
- Bayesian Test of Demographic Parity for Learning to Rank
  PH Kanani, H Mahmoudian, S Panda, JB Tristan, ML Wick
- Character-based attribute value extraction system
  PH Kanani, ML Wick, AC Pocock
  US Patent App. 14/700,683
- Named entity recognition and entity linking joint training
  PH Kanani, ML Wick, K Silverstein
  US Patent App. 15/168,309
- Multilingual embeddings for natural language processing
  ML Wick, PH Kanani, AC Pocock
  US Patent 9,779,085
- Multi-dimensional approach to agent assignment
  D Allison, D Gulsen, VCW Chan, AC Pocock, P Kanani, D Greenberg
  US Patent App. 14/804,496

HONORS and AWARDS

- Accomplishments in Search & Mining Award, UMass Amherst Computer Science, 04/2011
- PAKDD'10 Best Student Paper, Runner Up, 06/2010
- IJCAI'07 Student Travel Award, 01/2007
- Verizon Rising Star Fellowship, 09/2006
- Second Position, IARPA KDD Challenge for Entity Resolution, 08/2005.

PROFESSIONAL ACTIVITIES

- **Area Chair/Senior Program Committee Member:** WiML 2016, SDM 2017, WiML 2019
- **Program Committee Member:** EMNLP'11, EMNLP-CoNLL'12, AKBC-WEKEX'12, WiML'12
- **Conference Reviewing:** NESCAI'08, ICML'09, NAACL-HLT'09, NESCAI'10, WSDM'10, NAACL'15, ICML'15, NIPS'15, ECML-PKDD'16, EMNLP'17, NIPS'17, KDD'18, EMNLP'18, ICML'18, NIPS'18, SDM'19, ICML'19, KDD'19, IJCAI'19, SocialNLP'19, NeurIPS'19, EMNLP'19, AAAI'19
- **Journal Reviewing:** INFORMS Journal on Computing (2012), ML Journal (2016), DMKD Journal (2016)
- **Senior Advisory Board**: Women in Machine Learning, premier organization promoting women in the international Machine Learning community.
- **Organizer:** WiML2012 (Women in Machine Learning workshop, co-located with NIPS 2012), Machine Learning and Friends Lunch at UMass Amherst (2005-2006)
- **Invitation-only Workshops:**
  o HIPstIR 2019: The hip "stick, sand, and paper" retreat on the future of information retrieval, 09/19
  o **DARPA ISAT** Toolkit of Algorithms for Machine Learning (TAMALE) Workshop, 03/18 Organized by a DARPA-sponsored study group, Information Science and Technology (ISAT) which, for 30+ years, has been a study group of academics and other researchers

interested in making sure that DARPA remains aware of emerging technologies which might be in the national interest. Recognized as one of the experts in the field

- o IBM Statistical Machine Learning and Its Applications (SMiLe) Workshop, 09/09
  Invited selected researchers and graduate students with promising research

INVITED TALKS

- "Machine Learning in the Wild: My research adventures at Oracle Labs", UMass CS Women's event, Amherst, India, 11/2018
- "Resource-bounded Information Acquisition and Learning", Yahoo Research Lab, Bangalore, India, 06/2010
- "Resource-bounded Information Acquisition and Learning", Natural Language Systems Group, IBM T. J. Watson Research Center, Yorktown Heights, 05/2008
- "Improving Author Coreference by Resource-bounded Information Gathering from the Web". Indian Institute of Technology, Mumbai, India, 01/2007

TECHNICAL SKILLS

- Languages/Scripting: Java, C/C++, Perl, Python, Scala, SQL, Lisp, Prolog, JavaScript
- Database/Data Access/Analysis: MATLAB, R, SPSS, Oracle 8i, MySQL, JDBC
- Machine Learning/NLP toolkits: Sci-kit Learn, Factorie, Mallet, Weka, Bayes Net toolbox, AEther (Microsoft proprietary tool)

GRADUATE COURSEWORK

UMass:
- Advanced Machine Learning
- Machine Learning
- Advanced Algorithms
- Theory of Computation
- Research Methods

Coursera:
- Probabilistic Graphical Models

NYU:
- Advanced Lab (Advanced Topics in AI)
- Natural Language Processing
- Machine Learning
- Artificial Intelligence
- Web Search Engines
- Mathematical Statistics
- Advanced Database Systems

TEACHING EXPERIENCE

- **Artificial Intelligence**   **UMass Amherst**
  Teaching Assistant (09/2010 – 12/2011)

- **Introduction to Scientific Teaching**   **UMass Amherst**
  Participated in the seminar course and designed a microteaching module (01/2009 – 05/2009)

- **Introduction to Natural Language Processing**   **UMass Amherst**
  Guest lecture on Information Extraction (12/2007)

- **Introduction to problem solving using computers – Honors section**   **UMass Amherst**
  Teaching Assistant (08/2006 – 12/2006)
  Helped design the course material, held office hours, graded homework and tests

- **Computational Linguistics**   **UMass Amherst**
  Teaching Assistant
  Held office hours, graded homework and tests

- **Foreign Academic Consultancy and Training**   **Mumbai, India**

Tutored quantitative and analytical skills to 150 students appearing for the GRE, GMAT and SAT tests. Also assisted in creating the course material. (10/2000 – 09/2001)

ACADEMIC PROJECTS

- Entity Resolution with Spectral methods on Hypergraphs
- Learning to predict Branches. Applied Machine Learning to Branch Prediction.
- Applying spectral clustering to the problem of author coreference.
- Applied data mining techniques on past bond market performance data in order to evaluate the existing financial model for a leading investment analysis firm.
- FindGuru: an information extraction system to mine faculty and students homepages. Used natural language processing to create an interactive database of people in the academia, their research interests and contact information.
- A term paper titled "Knowledge Sharing: from Bibliography to Autonomous Citation Indexing and beyond".
- Implemented a system to extract melodies used for matching against user queries.
- Developed a system for indexing and querying efficiently from web documents.
- Implemented a Naive Bayes language classifier.
- Created a subject focused web crawler.
- Experiments on comparing the relative performance of text mining algorithms.
- Simulated concurrency control and recovery of replicated data.
- Designed a system for planning resource allocation in a manufacturing unit.

VOLUNTEER

- Regularly give talks to local middle and high school children about Machine Learning
- Invited on the panel for AI4All to talk to young women about careers in AI
- Invited on the panel for industry mentorship class at UMass, Amherst
- Invited to talk about career growth at the UMass CS Women's group
- Events coordinator for the Indian Students Association, organizing and participating in major cultural events during 2005.
- Active participant for the annual OISS cultural festival during 2002 and 2003.
- Head of 'Fine Arts Events' in the inter-collegiate cultural festival, Horizons 2000.

LANGUAGES KNOWN

- English, Gujarati, Hindi, Marathi.

INTERESTS

- Art, Music, Classical and Folk Dance, Poetry, Literature, Cooking, Knitting, Traveling.

REFERENCES

- Available upon request.