Bipasha Sen

bipasha.sen3195@gmail.com | +918108980291
202 Block - 8, Hill Ridge Springs Gachibowli Hyderabad Telangana

Education

University of Mumbai — Mumbai, Maharashtra
Bachelor of Technology in Computer Science; GPA: 8.14/10.0 — *Aug 2012 - May 2016*

Thesis: SmartShuffle: This is what I wanted to hear!

Relevant Coursework: Applied Mathematics, Data Structures, Analysis of Algorithms, Artificial Intelligence, Machine Learning

KV ONGC — Panvel, Maharashtra
12th CBSE; 82% — *Apr 2011 - Mar 2012*

KV ONGC — Panvel, Maharashtra
10th CBSE; 91% — *Apr 2009 - Mar 2010*

Skills

- **Languages**: C#, Python, C++ Front-End (jQuery, HTML, CSS), Query-Languages (USQL, Spark.net, MySQL), PHP, Java
- **Technologies**: Deep Learning, Full Stack, Object Oriented Design
- **Designing Tools**: Adobe Premiere Pro, Adobe Photoshop, Adobe After Effects, Sony Vegas Pro

Experience

Microsoft India R&D — Hyderabad, Telangana
Data Scientist - Microsoft Search Assistant and Intelligence (MSAI) — *July 2016 - Present*

- **Learning to Rank on Clicked Logs for Contextual Suggestions on Outlook Emails**:
  - Developing deep learning models for contextual suggestions on emails attachments using click logs. Defining an evaluation metric and a loss function to optimize on. Challenges: Weakly labeled dataset, position bias, selection bias, noise.
- **Meetings Insights - Intelligent Recommendations on Outlook Meetings**: This scenario impacts over 100million users per month.
  - Developed an end-to-end training pipeline for iterative model development at scale on a dataset of 50 million meetings. Challenges: Class-imbalance, Weakly labeled data, Bias.
- **Inference of Logical Entities called Trips on Machine-to-Human emails on Outlook**: When a user books an end-to-end trip, multiple bookings are made (flights, hotels, rental cars e.t.c) which in turn results in multiple machine-generated reservation emails in their inbox. Such reservation emails that appear to form a part of single logical unit are clubbed together to form Trips.
  - Responsible for the subjective analysis of how a trip is perceived by individuals given a set of elements namely flights, hotels etc. Development of a merging logic to form an end-to-end meaningful trip. Challenges: round trips, intersecting trips etc.
- **Editable-summarization of information extracted on Machine-to-human emails on Outlook**: Key information extracted from machine-generated emails, such as bill due, reservation emails etc are presented as a summarized view called cards. This features enables the users to rectify and edit any incorrect extraction such that the edited information is propagated to future similar emails.
  - I owned the implementation of editable-summarization for Invoice, Service Appointments and Logding Reservation scenarios.
- **Automated pipeline for validation of information extracted on Machine-to-human emails on Outlook**: Automated information extraction on emails called for manual validation for evaluating the quality of extraction. The process was automated by employing heuristics and anomaly detection techniques.
  - I owned the Flight Reservation and Lodging Reservation scenario.
- **Scalable non-template based approach for information extraction on Machine-to-human emails on Outlook using PROSE**: Microsoft Program Synthesis using Examples (PROSE) is a program synthesis framework for domain-specific languages (DSLs).
  - Responsible for extending and integrating PROSE for scalable non-template based information extraction on structured and non-structured natural language emails.

Microsoft India R&D — Hyderabad, India
Research Intern — *Dec 2015 - Feb 2016*

- **The bot will help you shop!**: A virtual shop assistant responsible for proactively engaging the users and assist them towards task completion. From a a set of curated questions, the agent needs to learn the most efficient order of questions to ask the users in order to maximize the engagement and win rate
  - Responsible for integrating Multi World Testing (Microsoft Research) with the bot by defining the policy and the reward functions.

I demoed this project to David Ku (former CVP and CTO of Microsoft AI+R) and was highly appreciated.

PROJECTS

- **Short-Term Context Based Fast Multilingual Acoustic Model for Low Resource Langauges, paper** : In this work, an architecture based on short term contextual temporal features learned on convolutional neural networks (CNNs) with a non-sequential discriminative network is proposed to reduce the inference time by a factor of 26.
- **Pose Based Action Recognition using Hierarchical Bidirectional Long Short Term Memory Network, paper** : In this work, we propose an end-to-end pipeline for the task of human action recognition on video sequences using 2D joint trajectories estimated from a pose estimation framework. We use a Hierarchical Bidirectional Long Short Term Memory Network (HBLSTM) to model the spatio-temporal dependencies of the motion by fusing the pose based joint trajectories in a part based hierarchical fashion.
- **Sentence Modelling for Contextual Meeting Segmentation, paper** : We propose a novel technique of contextual meeting segmentation for the task of meeting summarization. We use pointer mechanism to extract the related sentences from a meeting transcription without assuming that the sentences are consecutive in nature.
- **SmartShuffle: This is what I wanted to hear!**: A prediction model that could predict the songs that a user would want to play next without requiring his intervention based on the current history.
  - The model works by detecting similarity between songs to learn a predictive model without using metadata such as sound-wave, song-name, genre etc. The idea was that similarity between two songs is quite subjective and differs heavily between individuals when the set of available songs is limited. Two songs with completely different meta properties can be perceived similar by an individual.
- **Anterior Segment Imaging - Collaboration with MIT Media Lab's ReDx**: Corneal anomaly detection to distinguish between a healthy and an unhealthy cornea with abnormalities on the 3D reconstructed image captured from a prototype for a low-cost, wearable solid-state device with no moving parts as an alternative to the existing bulky, expensive ophthalmic slit lamp.
- **Cloud-Based Group-Oriented file sharing network - theBhaad.com, video**: Self-motivated project undertaken in sophomore year of undergraduate degree with the motivation to build a better system to share assignments and documents with peers.
  - The development of the portal was undertaken in order to have an one stop virtual environment that compliments the real environment interactions between students and professors. The idea was build a file-sharing *network* instead of just a portal.
  - I developed the portal end-to-end from scratch with any operating system like user-interface for easy operation with an advanced search features across groups, contacts, within groups, implemented auto-sorting techniques that sorted the documents by their importance at any given point of time, discussion forum, request and push-notification features.

  This was extensively used by my undergraduate institution at a time having on an average of 5000 active users per month

ADDITIONAL EXPERIENCE & ACHIEVEMENTS

- Invited talk on *Short Term Context Based Multilingual Acoustic Model for Low Resource Languages* in Microsoft's Machine Learning and Data Science conference.
- *3rd place in Microsoft One Week Challenge*, an internal worldwide hackathon with 3k+ participants.
- Awarded *Best Student of the year '15-16* by my undergraduate university
- Awarded *Best Entrepreneur* for my work on theBhaad.com by my undergraduate university.
- *Winner* of HackerRank's Magic Lines September '15
- *126th place* in TCS CodeVita '15 Round 2 out of 19800 hackers. Received on the spot employment offer from TCS.
- *197th place* in HackerRank's Software Challenge '14 out of 3000 hackers, *29th place* in Hack 101 September Challenge '14
- Extra curricular: Vocalist, Guitarist, Composer, Painter, Sketcher, YouTuber, Traveler.