DATTA MEGHE

INSTITUTE OF HIGHER EDUCATION & RESEARCH

(DEEMED TO BE UNIVERSITY)

LEARN. LEAD.

DATTA MEGHE
INSTITUTE OF HIGHER
EDUCATION & RESEARCH
(DEEMED TO BE UNIVERSITY)

LEARN. LEAD.

Faculty of Engineering and Technology

Department of Artificial Intelligence and Data Science

DATTA MEGHE INSTITUTE OF HIGHER EDUCATION & RESEARCH
SAWANGI (MEGHE), WARDHA

FACULTY OF ENGINEERING AND TECHNOLOGY, SAWANGI (MEGHE) , WARDHA

MOOC Course
on
Coursera Courses & NPTEL - Natural Language Processing

NAME OF FACULTY: DR. UTKARSHA PACHARNEY

NAME OF STUDENTS: BHAIRAVI NARENDRA REWATKAR

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

DATE :03/03/2025

Introduction to Statistics & Data Analysis in Public Health

- This course introduces fundamental statistical techniques such as descriptive statistics (mean, median, standard deviation) and inferential statistics (regression analysis, hypothesis testing) used for analyzing public health data.
- It covers various graphical tools like histograms, box plots, scatter plots, and heatmaps to represent healthcare data effectively.
- The course also explains different study designs, including cross-sectional, case-control, and cohort studies.
- Finally, the course explores predictive modeling techniques used in healthcare, such as logistic regression for disease classification and time-series analysis for outbreak forecasting.

Data Science in Stratified Healthcare and Precision Medicine

- The course introduces the concept of stratified healthcare, where patients are categorized into subgroups based on genetic, clinical, and lifestyle factors to better tailor treatment approaches.
- It explores the role of machine learning algorithms like decision trees, random forests, and neural networks in predicting disease risks and treatment responses within these subgroups.
- The course covers how genomic sequencing and biomarker identification help in precision medicine, enabling personalized treatment plans for conditions like cancer and cardiovascular diseases.
- AI-driven analytics have been successfully applied in healthcare, with examples including AI-based cancer treatment recommendations, personalized drug prescriptions for chronic diseases, and predictive models for early-stage disease detection.

Clinical Decision Support Systems (CDSS)

- AI-driven CDSS utilize deep learning and natural language processing (NLP) to analyze patient symptoms, lab reports, and medical histories.
- These systems suggest potential diagnoses, recommend treatments, and even predict disease progression, helping doctors enhance decision-making and improve patient outcomes.
- CDSS are designed to seamlessly integrate with Electronic Health Records (EHRs) and hospital management systems to provide real-time alerts and personalized treatment plans.
- A major challenge in CDSS adoption is interoperability, as different hospitals use diverse EHR systems with varying data formats.

Google Prompting Essentials

- It explores how Large Language Models (LLMs) process user inputs and generate relevant responses, emphasizing accuracy and coherence.
- It covers zero-shot, few-shot, and chain-of-thought prompting, improving complex problem-solving.
- It discusses methods to structure prompts effectively, including clarity, specificity, and context enhancement to maximize AI response quality.
- It highlights how AI prompting is transforming industries such as healthcare (AI-driven diagnostics), finance (automated reporting), customer support (AI chatbots), e-commerce (personalized recommendations), and content creation (marketing copy and blog writing automation).

Course Overview (NPTEL - Natural Language Processing)

- This course led by Prof. Pawan Goyal at IIT Kharagpur, provides a structured approach to Natural Language Processing (NLP), covering both fundamental linguistic principles and advanced computational techniques.

- The course is backed by tech giants like Microsoft Research, Google, Adobe, and Amazon, ensuring exposure to cutting-edge industrial applications.

- Covers core linguistic concepts like syntax, semantics, and morphology, along with machine learning and deep learning approaches for text classification, information retrieval, and entity recognition.

- The course emphasizes practical learning through coding exercises, NLP case studies, and dataset analysis.

- Explores speech recognition, machine translation, chatbot development, and automated text summarization. Students gain expertise in deploying NLP in financial services (automated trading insights), healthcare (clinical text analysis), and customer support (intelligent chatbots like ChatGPT and Bard).

Week 1-4 :

- Basic Text Processing (Week 1) – Introduces essential text preprocessing techniques like text normalization (lowercasing, punctuation removal), tokenization (splitting text into words or subwords), stopword removal (eliminating common words like "the" and "is"), and stemming/lemmatization (reducing words to their root form for uniformity in analysis).

- Spelling Correction & Language Modeling (Week 2) – Discusses techniques for identifying and correcting spelling errors using algorithms like edit distance (Levenshtein Distance) and phonetic matching. Language modeling methods, including n-gram models and neural networks, are introduced to predict word sequences and improve text understanding.

- Advanced Smoothing & POS Tagging (Week 3) – Explores different smoothing techniques like Laplace smoothing, Good-Turing estimation, and Kneser-Ney smoothing to address data sparsity in language models. POS tagging assigns grammatical categories (e.g., noun, verb, adjective) to words using rule-based, statistical, and deep-learning methods.

- Sequential Tagging Models (Week 4) – Covers probabilistic models like Maximum Entropy Markov Models (MaxEnt) and Conditional Random Fields (CRF) for sequence labeling tasks. These models are used in Named Entity Recognition (NER), POS tagging, and chunking to identify patterns in text sequences.

Week 5-7 :

- Syntax – Constituency Parsing (Week 5) – Focuses on breaking down sentences into hierarchical tree structures based on phrase structure rules. Algorithms like the CYK (Cocke–Younger–Kasami) parser and probabilistic context-free grammars (PCFG) are used for parsing sentence structures.

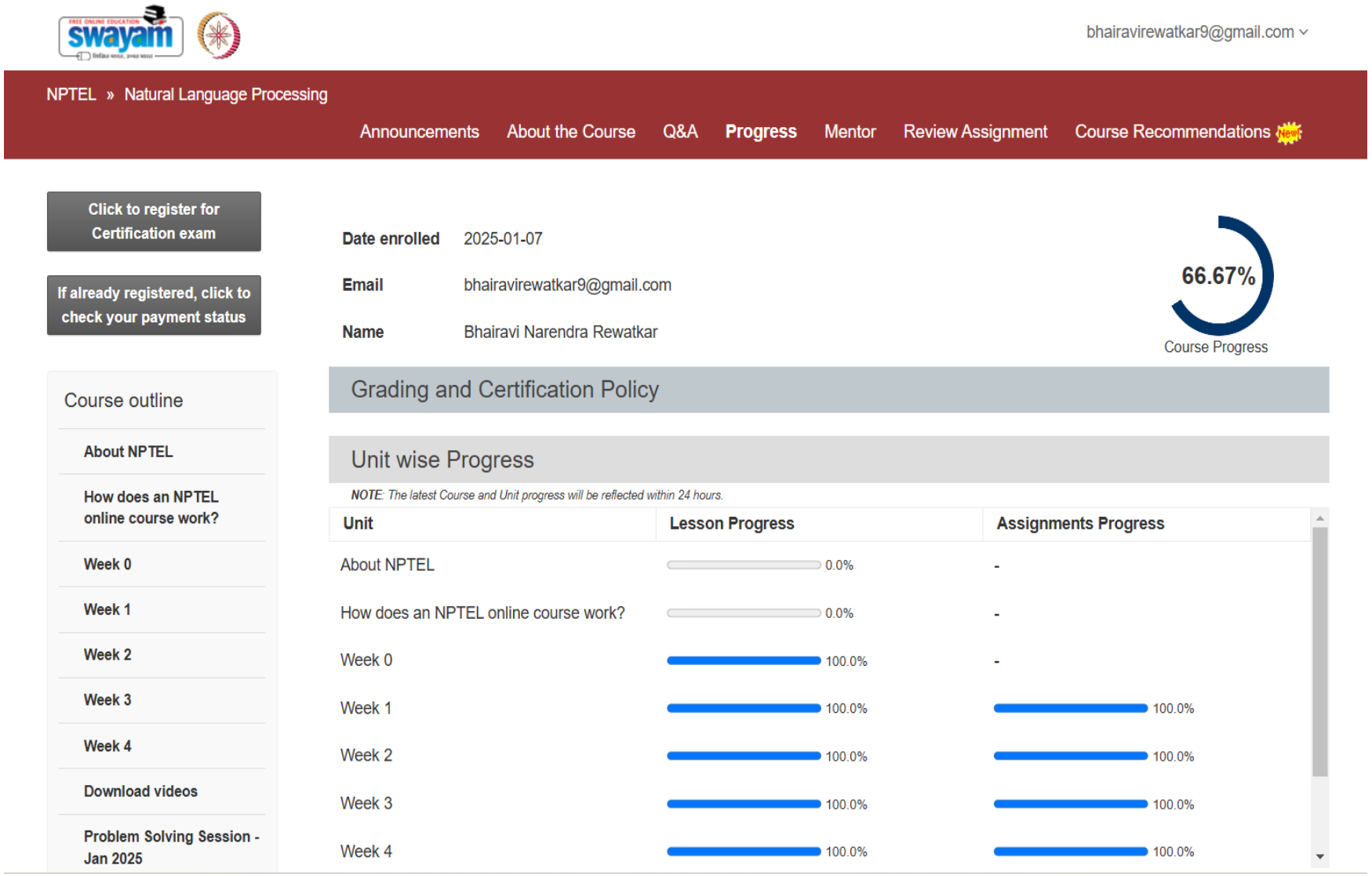
- Dependency Parsing (Week 6) – Examines how words in a sentence depend on each other using directed graphs. Dependency parsing helps in syntactic analysis by defining relationships such as subject-verb-object. Dependency parsing algorithms include the Eisner and MaltParser methods.

- Distributional Semantics (Week 7) – Introduces the concept that words with similar meanings occur in similar contexts. Techniques like Word2Vec, GloVe, and FastText use vector embeddings to represent words in a multi-dimensional space, allowing NLP models to capture semantic relationships between words.

Conclusion

- Gained expertise in AI applications for healthcare, precision medicine, and clinical decision-making, enabling data-driven insights for better patient outcomes.
- Explored machine learning models, NLP techniques, and decision support systems used in disease prediction, treatment optimization, and automated diagnostics.
- Developed proficiency in statistical analysis, data visualization, text processing, and deep learning to handle structured and unstructured healthcare data.
- Learned about AI bias, patient data privacy, and ethical concerns in deploying clinical decision support systems and AI-powered medical applications.
- Identified potential career paths in AI research, healthcare informatics, automation, and NLP advancements, paving the way for innovation in personalized medicine and medical AI.

NPTEL Course Progress



References

- https://coursera.org/verify/F7K4JMWBT438
- https://coursera.org/verify/K8W42YRC4BKH
- https://coursera.org/verify/I2MO3B5Z8KCO
- https://coursera.org/verify/HQNEP71H06ZR
- https://onlinecourses.nptel.ac.in/noc25_cs51/preview
- Dan Jurafsky and James Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall, Second Edition, 2009.
- Some draft chapters of the third edition are available online: https://web.stanford.edu/~jurafsky/slp3/

  Chris Manning and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA: May 1999.

Thank You...