

## CSCE 590-1: Trusted AI

# Lecture 16: Unstructured Text – Supervised ML

---

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

14<sup>TH</sup> OCT, 2021

***Carolinian Creed: “I will practice personal and academic integrity.”***

# Organization of Lectures 16

---

- Introduction Segment
  - Recap of Lecture 15
- Main Segment
  - Classifying with Textual data
  - Important considerations
- Concluding Segment
  - About next lecture – Lecture 17
  - Ask me anything

# Introductory Segment

---

# Schedule Snapshot



Sep 28 (Tu)	Review: AI Fairness, Project presentations, Discussion	Quiz 2
Sep 30 (Th)	AI - Unstructured (Text): Processing and Representation	
Oct 5 (Tu)	AI - Unstructured (Text): Common NLP Tasks	Mid-sem Project Review
Oct 7 (Th)	FALL BREAK	NO Classes, Course Midpoint
Oct 12 (Tu)	AI – Unstructured (Text): Analysis – NLP Tasks	
Oct 14 (Th)	AI – Unstructured (Text): Analysis – Supervised ML	
Oct 19 (Tu)	Invited Guest – AI - Supervised ML: External Talk/ AI Explanation Methods (AIX)	10 am EST
Oct 21 (Th)	Invited Guest – AI - Supervised ML: External Talk/ Working Session on AIX360	10 am EST
Oct 26 (Tu)	AI - Unstructured (Text): Analysis – Supervised ML – Trust Issues	Quiz 3
Oct 28 (Th)	Review: project presentations, Discussion	
Nov 2 (Tu)	AI - Unstructured (Text): Analysis – Supervised ML – Trust Issues	
Nov 4 (Th)	AI - Unstructured (Text): Analysis – Supervised ML – Mitigation Methods	
Nov 9 (Tu)	AI - Unstructured (Text): Analysis – Supervised ML – Explanation Methods	
Nov 11 (Th)	Trust: Data Privacy Trust: AI Testing	

# Recap of Lecture 15

---

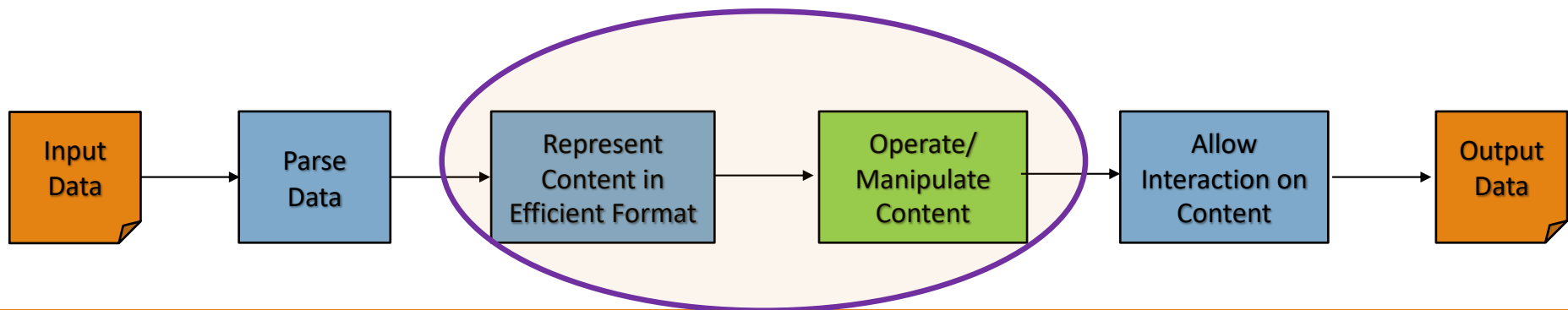
- Looked at more common NLP Tasks
  - Sentiment Detection
  - Question Answering
  - Summarization
- Saw working of pretrained embedding on some NLP Tasks
  - Using transformers

A variety of methods

- **Text similarity**
- Event Extraction
- **Sentiment detection**
- **Question Answering**
- **Summarization**
- Machine translation
- Natural Language Interface to Databases
- Natural Language Generation

# Main Segment

---



# Supervise Machine Learning with Text

---

# Concepts

---

- **Input data:** data available
  - **Training data:** used for training a learning algorithm and get a model
    - [Optional] **Validation data:** used to tune parameters
  - **Test data:** used to test a learning model
- **Prediction problem**
  - Learning value of a continuous variable
- **Classification problem**
  - Separating data into classes (also called labels, categorical types)
  - One of the attributes is the class label we are trying to learn
  - Class label is the **supervision**
- **Clustering problem**
  - We are trying to learn grouping of data
  - There is no attribute indicating membership in the groups (hence, **unsupervised**)

Reference: <https://machinelearningmastery.com/difference-test-validation-datasets/>  
<https://www2.seas.gwu.edu/~bell/csci243/lectures/classification.pdf>



# Classification, for What? (Applications)

---

- Binary
  - Fake v/s real content
  - Spam v/s non-spam
  - Positive v/s negative
  - Toxic v/s non-toxic
- Multi-class
  - Categories of news
  - ...

# Metrics: Accuracy, Precision, Recall

Actual Class	Predicted class		
		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

**Accuracy** =  
$$\frac{(TP+TN)}{(TP+FP+FN+TN)}$$

**Precision** =  
$$\frac{(TP)}{(TP+FP)}$$

**Recall** =  
$$\frac{(TP)}{(TP+FN)}$$

**F1 Score: Harmonic Mean**  
$$1/F1 = 1/Precision + 1/Recall$$

$$F1 = \frac{2 * (Recall * Precision)}{(Recall + Precision)}$$

# Criteria for Comparing Classification Methods

---

## Effectiveness

- Goodness of output: predictive accuracy, recall, F1
- Interpretability: providing insight
- Robustness: handling noisy data

## Efficiency

- Speed
- Scalability: large volume of data

**SOURCE:** Data Mining: Concepts and Techniques, by Jiawei Han and Micheline Kamber

# Common Steps

---

- Data processing
  - Acquire data
  - Prepare data
    - NLP methods to clean data and convert to efficient format
    - NLP tasks based on application
- Learning process
  - Training algorithm, procedures
  - Evaluation metrics
  - Optimizing learning procedure
- Communicating results/ Trust issues

# Code Example – Fake v/s Real News

---

**Sample code:**

<https://github.com/biplav-s/course-tai/blob/main/sample-code/l13-l16-supervised-text/l16-Text%20Classification.ipynb>

# Multi-Class / Categories

---

**Sample code:**

<https://github.com/biplav-s/course-tai/blob/main/sample-code/l13-l16-supervised-text/l16-text%20classification%20groups.ipynb>

# Class Discussion

---

- Review BBC category classification code (kernels) using pre-trained models / embeddings
- What NLP steps do they
  - Avoid with pre-trained models
  - Do additionally

# Practical Considerations

---

- Perspective based on review of Kaggle tasks and submissions
  - <https://neptune.ai/blog/text-classification-tips-and-tricks-kaggle-competitions>
- Understanding data and purpose of classification are key



# Explanation with Text

---

LIME: <https://marcotcr.github.io/lime/tutorials/Lime%20-%20basic%20usage%2C%20two%20class%20case.html>

# Concluding Segment

---

# Lecture 16: Concluding Comments

---

- We looked at Supervised ML with text
- Types
  - Binary
  - Multiple types/ categories
- Explanation method still very relevant for acceptance of decisions

# About Next Lecture – Lecture 17

---

# Schedule Snapshot



Sep 28 (Tu)	Review: AI Fairness, Project presentations, Discussion	Quiz 2
Sep 30 (Th)	AI - Unstructured (Text): Processing and Representation	
Oct 5 (Tu)	AI - Unstructured (Text): Common NLP Tasks	Mid-sem Project Review
Oct 7 (Th)	FALL BREAK	NO Classes, Course Midpoint
Oct 12 (Tu)	AI – Unstructured (Text): Analysis – Supervised ML	
Oct 14 (Th)	AI – Unstructured (Text): Analysis – Supervised ML	
Oct 19 (Tu)	Invited Guest – AI - Supervised ML: External Talk/ AI Explanation Methods (AIX)	10 am EST
Oct 21 (Th)	Invited Guest – AI - Supervised ML: External Talk/ Working Session on AIX360	10 am EST
Oct 26 (Tu)	AI - Unstructured (Text): Analysis – Supervised ML – Trust Issues	Quiz 3
Oct 28 (Th)	Review: project presentations, Discussion	
Nov 2 (Tu)	AI - Unstructured (Text): Analysis – Supervised ML – Trust Issues	
Nov 4 (Th)	AI - Unstructured (Text): Analysis – Supervised ML – Mitigation Methods	
Nov 9 (Tu)	AI - Unstructured (Text): Analysis – Supervised ML – Explanation Methods	
Nov 11 (Th)	Trust: Data Privacy Trust: AI Testing	

# Schedule Snapshot



Sep 28 (Tu)	<b>Review: AI Fairness, Project presentations, Discussion</b>	Quiz 2
Sep 30 (Th)	AI - Unstructured (Text): Processing and Representation	
Oct 5 (Tu)	AI - Unstructured (Text): Common NLP Tasks	Mid-sem Project Review
Oct 7 (Th)	FALL BREAK	<b>NO Classes, Course Midpoint</b>
Oct 12 (Tu)	AI – Unstructured (Text): Analysis – NLP Tasks	
Oct 14 (Th)	AI – Unstructured (Text): Analysis – Supervised ML	
Oct 19 (Tu)	<b>Invited Guest</b> – AI - Supervised ML: External Talk/ AI Explanation Methods (AIX)	10 am EST
Oct 21 (Th)	<b>Invited Guest</b> – AI - Supervised ML: External Talk/ Working Session on AIX360	10 am EST
Oct 26 (Tu)	AI - Unstructured (Text): Analysis – Supervised ML – Trust Issues	Quiz 3
Oct 28 (Th)	<b>Review: project presentations, Discussion</b>	
Nov 2 (Tu)	AI - Unstructured (Text): Analysis – Supervised ML – Trust Issues	
Nov 4 (Th)	AI - Unstructured (Text): Analysis – Supervised ML – Mitigation Methods	
Nov 9 (Tu)	AI - Unstructured (Text): Analysis – Supervised ML – Explanation Methods	
Nov 11 (Th)	Trust: Data Privacy Trust: AI Testing	

# Lecture 17 and 18: Explanation Methods

---

- Explaining working of classification methods
- Working session with AIF 360

## About Speaker - 1

Oct 19 (Tu)	<b>Invited Guest</b> – AI - Supervised ML: External Talk/ AI Explanation Methods (AIX)	10 am EST
Oct 21 (Th)	<b>Invited Guest</b> – AI - Supervised ML: External Talk/ Working Session on AIX360	10 am EST

- **Diptikalan Saha**, IBM Research

- <https://researcher.watson.ibm.com/researcher/view.php?person=in-diptsaha>

Dr. Diptikalan Saha (Dipti) is a Senior Technical Staff Member and manager of Reliable AI team in Data&AI department of IBM Research at Bangalore. His research interest includes Artificial Intelligence, Natural Language Processing, Knowledge representation, Program Analysis, Security, Software Debugging, Testing, Verification, and Programming Languages. received my Ph.D. degree in Computer Science from the State University of New York at Stony Brook. My advisors were Prof. C. R. Ramakrishnan and Prof. Scott A. Smolka. I received my B.E. degree in Computer Science and Engineering from Jadavpur University. His group's work on Bias in AI Systems is available through [AI OpenScale in IBM Cloud](#) as well as through open-source [AI Fairness 360](#).



## About Speaker - 2

Oct 19 (Tu)	<b>Invited Guest</b> – AI - Supervised ML: External Talk/ AI Explanation Methods (AIX)	10 am EST
Oct 21 (Th)	<b>Invited Guest</b> – AI - Supervised ML: External Talk/ Working Session on AIX360	10 am EST

- **Vijay Arya**, IBM Research
  - <https://researcher.watson.ibm.com/researcher/view.php?person=in-vijay.arya>

Vijay Arya is a senior researcher in IBM Research AI at the IBM India Research Lab where he works on problems related to Trusted AI. Vijay has 15 years of combined experience in research and software development. His research work spans Machine learning, Energy & smart grids, network measurements & modeling, wireless networks, algorithms, and optimization. His work has received outstanding technical achievement awards at IBM and has been deployed by power utilities in USA. Before joining IBM, Vijay worked as a researcher at National ICT Australia (NICTA) and received his PhD in Computer Science from INRIA, France, and a Masters from Indian Institute of Technology (IIT) Delhi. He has served on the program committees of IEEE, ACM, and IFIP conferences, he is a senior member of IEEE & ACM, and has more than 60 conference & journal publications and patents.