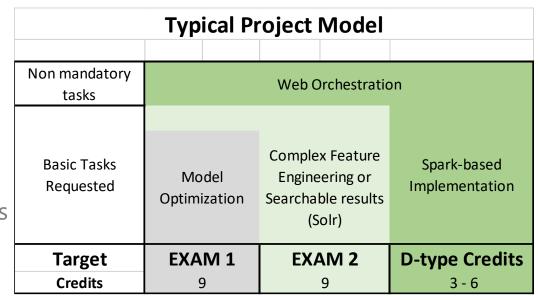# Project Introduction

WM&R a.a. 2018-19,

R. Basili

Project Types and Team Size

Types of Project & Activities

- Model optimization & ML Algorithm comparison

- Feature Engineering

- Image processing & ML for Object Recognition

- Integration of a Search Engine over classified entries

- Integration in a distributed Big Data architecture

Typical Project Model			
Non mandatory tasks	Web Orchestration		
Basic Tasks Requested	Model Optimization	Complex Feature Engineering or Searchable results (Solr)	Spark-based Implementation
Target	**EXAM 1**	**EXAM 2**	**D-type Credits**
Credits	9	9	3 - 6

Team size: from 1 to 3 depending on the types of project selected

Project Activities

Target tasks

- Regression, Classification, Sequence Labeling
- Object Recognition from images
- Text Generation (through RNNs)
- Adversarial Machine Learning (GAN)

Data Gathering

- Download, feature extraction and modeling

Model Optimization: Test different algoritms, different features and different parametrizations

Feature engineering: Extend data sets with new features from external sources, e.g. Wikipedia

Project Activities (2)

Searchable Results

- ML is used to enrich the input data (e.g. tweets)
  - ML for automatic metadata creation
- Indexing is applied to data (+metadata) through IR engines (e.g. Solr/Lucene)
- Querying through a dedicated front-end
- Ranking model based on source information plus enrichment metadata

Distributed Architecture

- The above ML and Search functionalities should be realised through Web services in a Big Data framework, e.g. SPark

Project Table (1)

Pr. Nr.	Project/Dataset	Link	Target Task	ML methods	#Subtasks	Comput. Compl.	XOR policy	Feature Engin.	Search	Team Size (min,max)
1	Hate Speech (Kaggle)	https://www.kaggle.com/vkrahul/twitter-hate-speech	Text Class	Classification	3	No	No	No	-	(1,1)
2								Yes	No	(2,2)
3								Yes	Yes	(2,3)
4	Semeval 2019: HateEval	https://competitions.codalab.org/competitions/19935	Text Class	Classification	3	No	No	No	No	(1,1)
5								Yes	No	(2,2)
6								Yes	Yes	(2,3)
7	Semeval 2019: Offense Recognition	https://competitions.codalab.org/competitions/20011	Text Class	Structured Classification	3	No	No	No	No	(1,1)
8								Yes	No	(2,2)
9								Yes	Yes	(2,3)
10	SemEval 2019 RumourEval	https://competitions.codalab.org/competitions/19938	Classification of Tree-structured conversations	Structured classification	3	No	No	No	No	(1,1)
11								Yes	No	(2,2)
12								Yes	Yes	(2,3)

Legend	na	Not Applicable	
	fac.	Facultative (non mandatory)	

Project Table (2)

Pr. Nr.	Project/Dataset	Link	Target Task	ML methods	#Subtasks	Comput. Compl.	XOR policy	Feature Engin.	Search	Team Size (min,max)
13	COCO Image Processing	http://cocodataset.org/#detection-2016	Image Proc. / Object F	Pattern Recognition	1	Yes	No	Yes	na	(2,2)
14	COCO Ita Captioning	http://cocodataset.org/#captions-2015	Image to Text transcri	Text Gen. (neural trans	1	Yes	No	Yes	na	(1,2)
15	COCO Eng Captioning	http://cocodataset.org/#captions-2015	Image to Text transcri	Text Gen. (neural trans	1	Yes	No	Yes	na	(1,2)
16	COCO image retrieval	http://cocodataset.org/#home	Image Caption Retriev	Learning to Rank	3	No	Yes	No	No	(1,1)
17							No	Yes	No	(1,2)
18							No	Yes	Yes	(1,3)
19	GAN on Face Generation		Image Generation	Generation/Classificati	1	Yes	Yes	No	na	(1,2)

Legend	na	Not Applicable	
	fac.	Facultative (non mandatory)	

Project Table (3)

Pr. Nr.	Project/Dataset	Link	Target Task	ML methods	#Subtasks	Comput. Compl.	XOR policy	Feature Engin.	Search	Team Size (min,max)
20	Bridgstone Hackaton (Driver		Driver detection	Stream Classification / Patt. Recognition	1	Yes	No	-	-	(2,2)
21	Bridgstone Hackaton (Tier consumption)		Tier Consumption Estim.	Regression	1	Yes	No	-	-	(2,2)
22	Bridgstone Hackaton (Driver Styles)		Drive Style Discovery	Clustering	1	Yes		-	-	(2,2)

Legend	na	Not Applicable	
	fac.	Facultative (non mandatory)	

Exam: the Roadmap

Written test
- Register for the second MidTerm and First Final
- Pass the written test & Get admission to the Project work

Select and Define the project
- Select the project topic/dataset
- Discuss team size and assess the needed activities with tutors

Close the practical tasks
- Validate your Lab exercise with D. Croce

Close the project
- Document your design and agree upon it with tutors
- Carry out all activities (with final demo and PPT presentation)

To do list

- Chose the project/dataset of interest

- Discuss the open aspects (e.g. team size or exams) with the prof

- Study the overall setting before the Summer (e.g. software tools, datasets, reference papers)

- Keep prof and tutors teachers updated along the project lifecycle with dedicated meetings

Contacts: {basili,croce}@info.uniroma2.it →