

A close-up photograph of a hand holding a blue pen, poised to write on a piece of paper. The hand is wearing a grey, textured sweater. The background is blurred, showing more of the paper and the pen.

1

# TEAM 43

**Abhijeet Kumar**

**Anurag Ghosh**

**Vatika Harlalka**

A decorative header featuring a large blue number '2' on the left. To its right is a horizontal band filled with various light blue line-art icons. These icons include a tag, puzzle pieces, a magnifying glass, a smartphone, a document, a speech bubble, a target, gears, a pie chart, an envelope, a thumbs up, a lightbulb, a clock, and a checkmark. The icons are arranged in a repeating pattern across the width of the header.

# 2

# AUTOMATED ESSAY GRADING

# 3

## Basic Idea

- ▶ Collection of Dataset from five sets of essays by American students from grade 7 to 10(150 to 550 words each).
- ▶ Extracting appropriate features (basis of concrete models) and filtering to reduce dimensionality.
- ▶ Various classifiers used to find grades having highest measure of Quadratic Weighted Kappa ie similarity with human grading scheme.

# 4

## Technique

- ▶ We have used various techniques like Support Vector Regression, SVM and Kernel Ridge Regression.
- ▶ We have also used Graph Diffusion techniques for analysis.
- ▶ Testing using k-fold cross validation.

### Citation:

Higgins, Derrick, Jill Burstein, Daniel Marcu, and Claudia Gentile. "Evaluating Multiple Aspects of Coherence in Student Essays." In HLT-NAACL, pp. 185-192. 2004

# 5

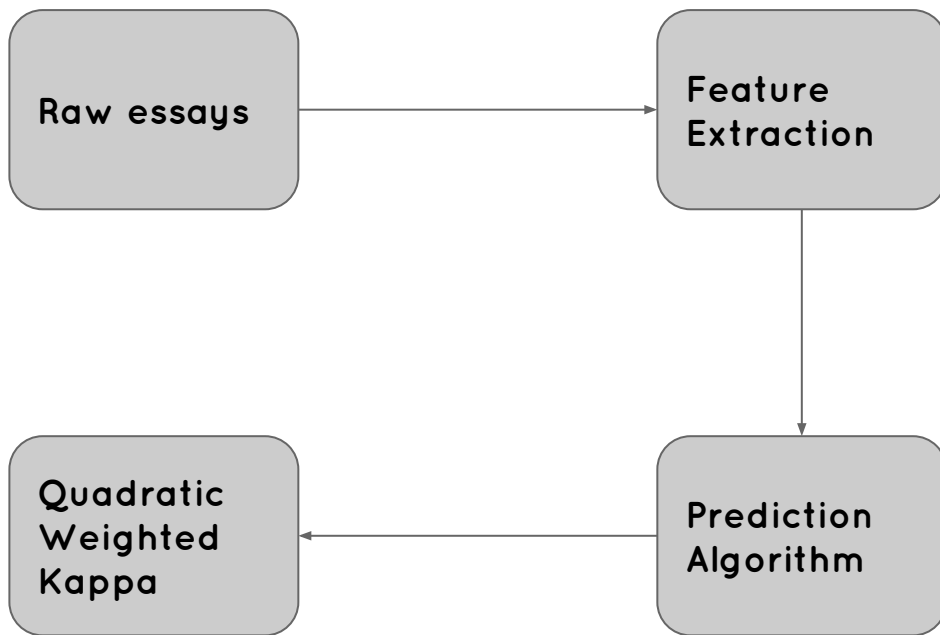
## Goal

The aim was to create a model which are able to make predictions closely matching with those by Human graders.

Find which features have greatest influence on the quality of the essay.

# 6

## Workflow



# 7

## Dataset

### Dataset

- ▶ Kaggle Dataset
- ▶ 5 sets with approximately 8000 essays ranging from 150-550 words each

# 8

## Feature Extraction

Features extracted from essays include:

- ▶ **Numerical features:** n-grams, average word counts, sentence counts, the number of words of different character lengths, number of sentences of different
- ▶ **Maturity features:** number of spelling errors, average age of acquisition of words, average beautiful word score
- ▶ **Semantic features:** Parts of Speech statistics like number of Nouns, Verbs, Adjectives, Bag of Words score, sentiment score, subjectivity score
- ▶ **Punctuation based features**

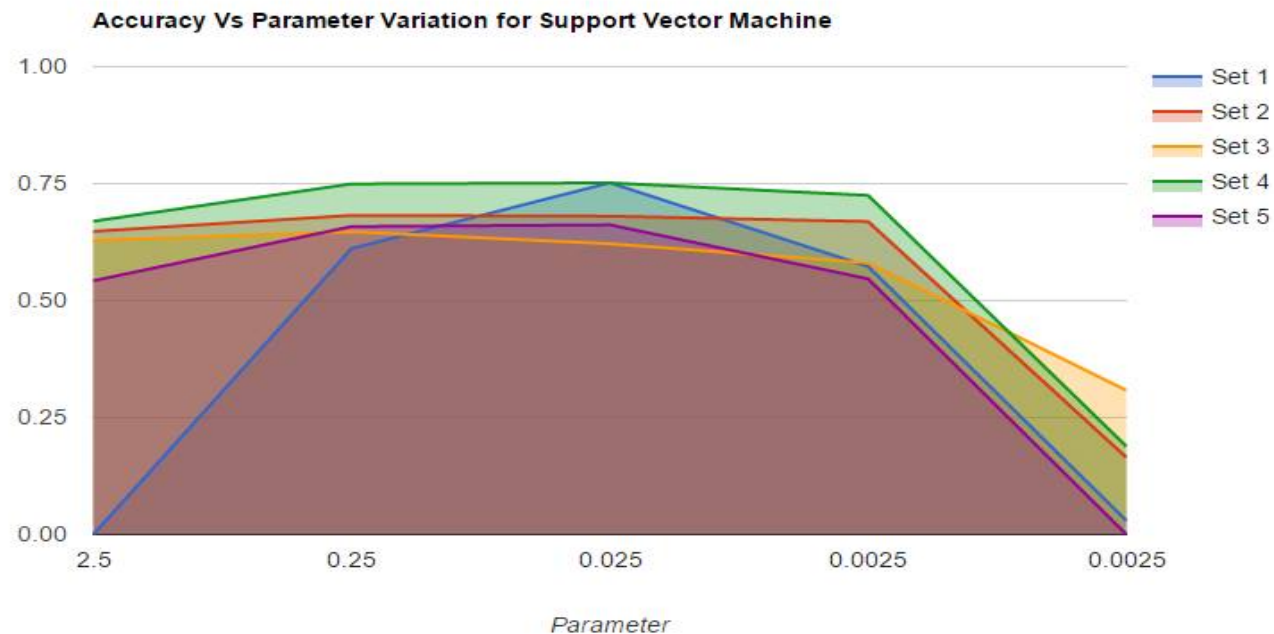


# 10

## Algorithm(1)

### Classification & Parameter Variation

#### ▷ SVM



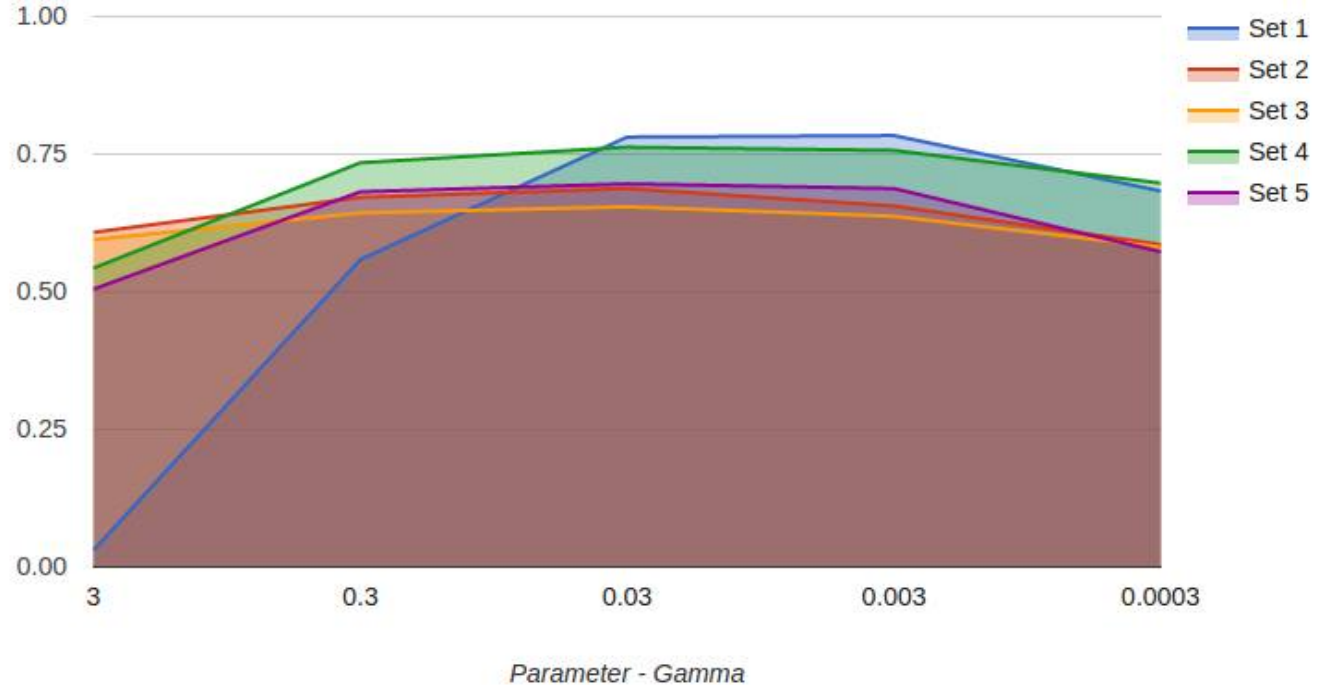
# 11

## Algorithm(2)

### Classification & Parameter Variation

#### ▷ SVR

Accuracy vs Parameter Variation for Support Vector Regression

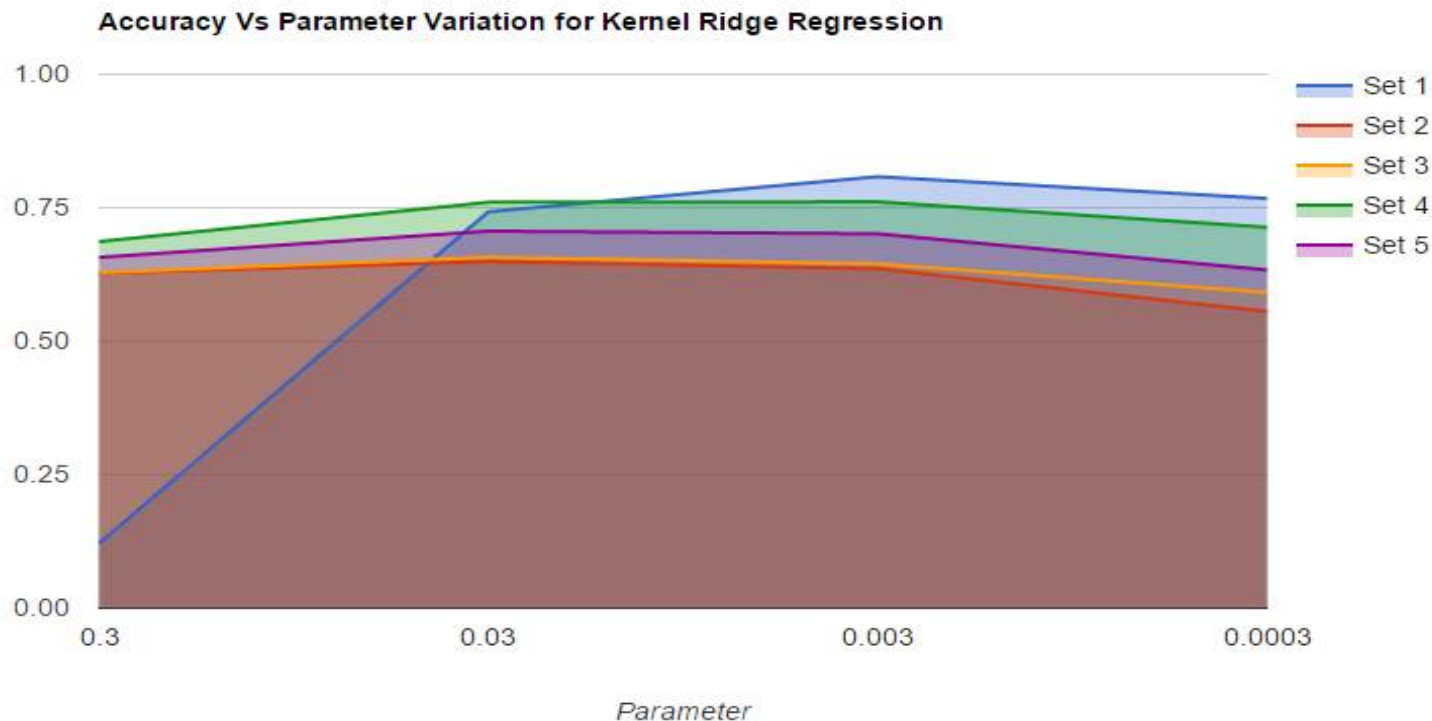


# 12

## Algorithm(3)

### Classification & Parameter Variation

#### ▸ Kernel Ridge Regression

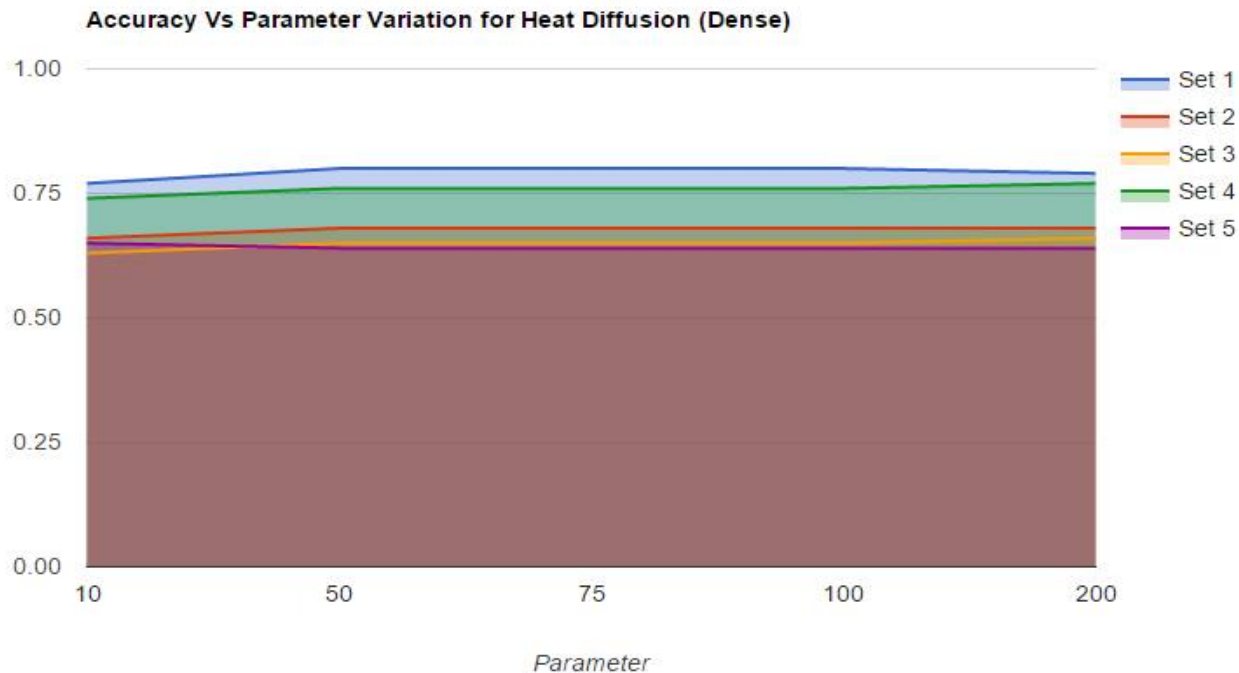


# 13

## Algorithm(4)

### Classification & Parameter Variation

#### ► Heat Diffusion



# 14

## Performance Measure

### Performance Measure

- ▶ Construct confusion matrix (C) by calculating number of times first grader gave grade i while second gave grade j.
- ▶ Find both  $p_o$  and  $p_e$  from C.
- ▶ Value can be between -1 and 1.
- ▶ Quadratic Weighted Kappa is given by

$$\kappa = (p_o - p_e) / (1 - p_e)$$

# 15

## Results

## Results

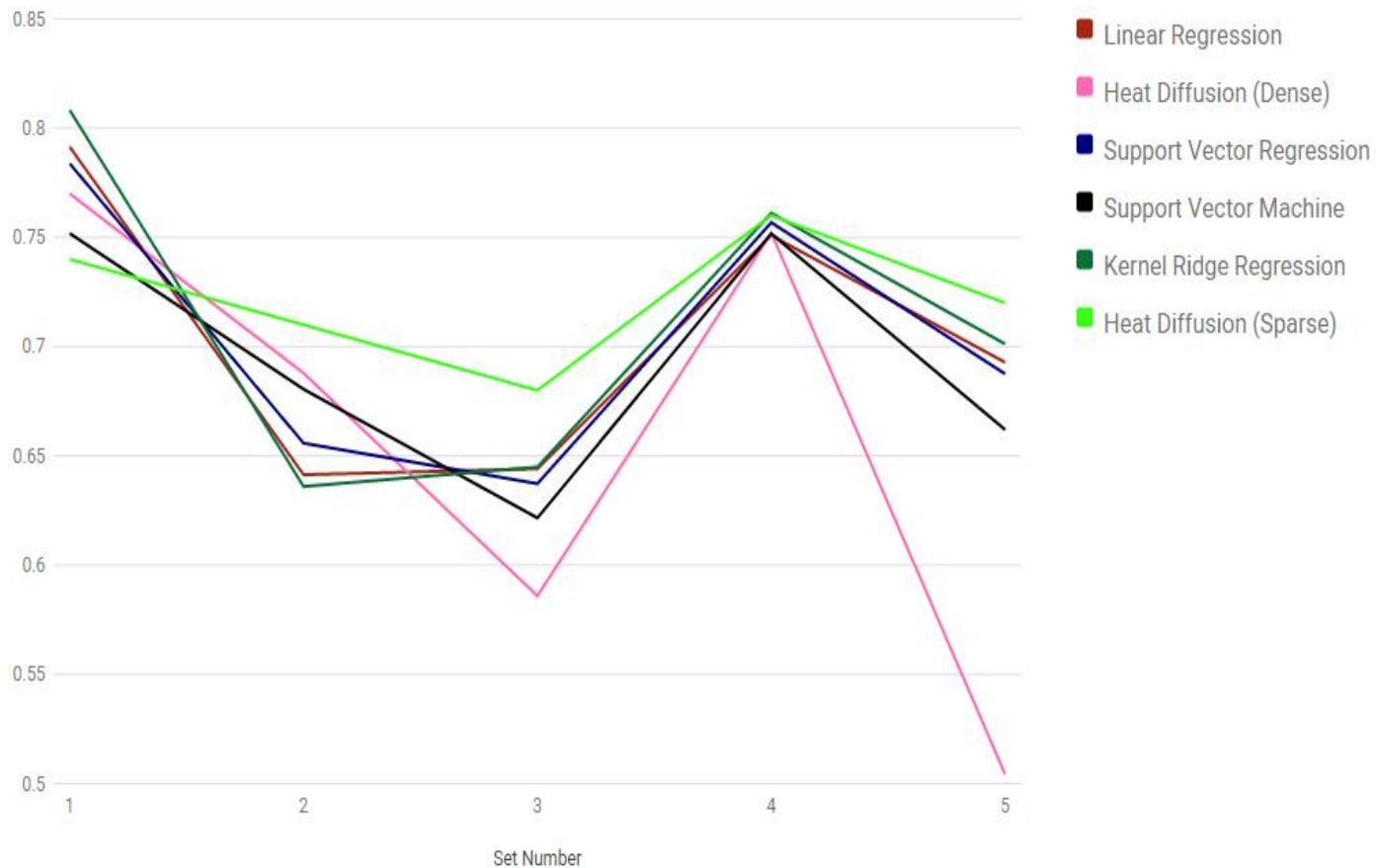
	Linear Regression	Heat Diffusion (Dense)	Support Vector Regression	Support Vector Machine	Kernel Ridge Regression	Heat Diffusion (Sparse)
1	0.79	0.77	0.78	0.75	0.81	0.74
2	0.69	0.69	0.66	0.68	0.64	0.71
3	0.64	0.59	0.64	0.62	0.64	0.64
4	0.75	0.75	0.76	0.75	0.76	0.76
5	0.69	0.50	0.69	0.66	0.70	0.72

# 16

## Result Comparison

Accuracy achieved using different Classifiers  
using Kappa Agreement Measure

Result



# 16

## Observations

### Notable Observations

- ▶ Heat Diffusion is significantly better with a proper graph construction, however robust to parameter variations.
- ▶ Linear Regression performs at par to SVM and SVR. Sometimes simple models work good enough.
- ▶ Kernel Ridge Regression performs slightly better than SVM and SVR with nearly equal variation.
- ▶ Heat Diffusion with Sparse Graph Construction shows the least variation with respect to other methods.



# 17

## FEATURE EXTRACTION

### Future Work

- ▶ **Improve feature set using domain expertise of Natural Language Processing.**
- ▶ **Bigger dataset needed to better represent classes which were severely underrepresented.**
- ▶ **Exploring other sparsification methods than currently employed in Heat Diffusion Implementation.**

# 18

## FEATURE EXTRACTION

### Acknowledgements

**We would like to thank Dr. Avinash Sharma, Ajitesh Gupta and Gaurav Mishra for their guidance to shape the project to its conclusion.**

A close-up photograph of a hand holding a blue pen, poised to write on a piece of paper. The hand is wearing a grey, textured sweater. The background is blurred, showing a desk and some papers.

# 19

# THANKS!

- ▶ [vatika.harilalka@research.iiit.ac.in](mailto:vatika.harilalka@research.iiit.ac.in)
- ▶ [halwai.abhijeet@gmail.com](mailto:halwai.abhijeet@gmail.com)
- ▶ [anurag.ghosh@research.iiit.ac.in](mailto:anurag.ghosh@research.iiit.ac.in)