

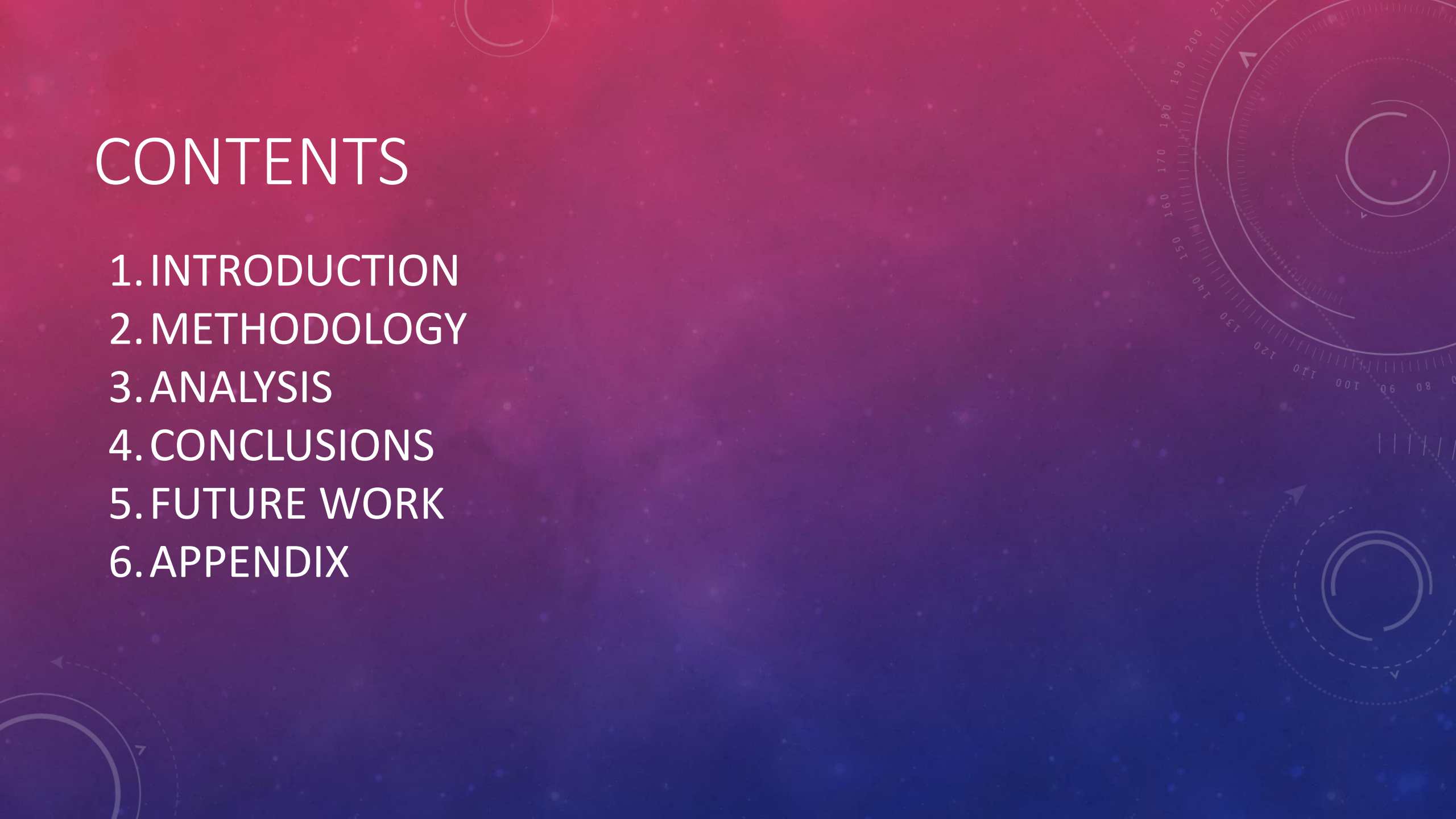


HALLMARKS OF ONLINE PRIVACY MINDERS

BY BURAK BASOGUL

CONTENTS

1. INTRODUCTION
2. METHODOLOGY
3. ANALYSIS
4. CONCLUSIONS
5. FUTURE WORK
6. APPENDIX



INTRODUCTION

- RULES, REGULATIONS & LAWS NOT CAUGHT UP WITH PRIVACY PROTECTION vs PERSONAL DATA COLLECTION
- INDIVIDUALS FEEL COMPELLED TO TAKE CONTROL OF THEIR ONLINE PRIVACY
- CAN WE IDENTIFY SOME CHARACTERISTICS OF PRIVACY MINDING INDIVIDUALS?

METHODOLOGY - 1

- DATA : AMERICAN TRENDS PANEL-49 BY PEW RESEARCH CENTER, JUN 2019
- 4,272 PARTICIPANTS, OVER 130 QUESTIONS w/CATEGORICAL RESPONSES
- SELECTED 87 FEATURES, TARGET INDIVIDUALS WHO FOLLOW PRIVACY NEWS “VERY CLOSELY”
- IMBALANCED TARGET DATA

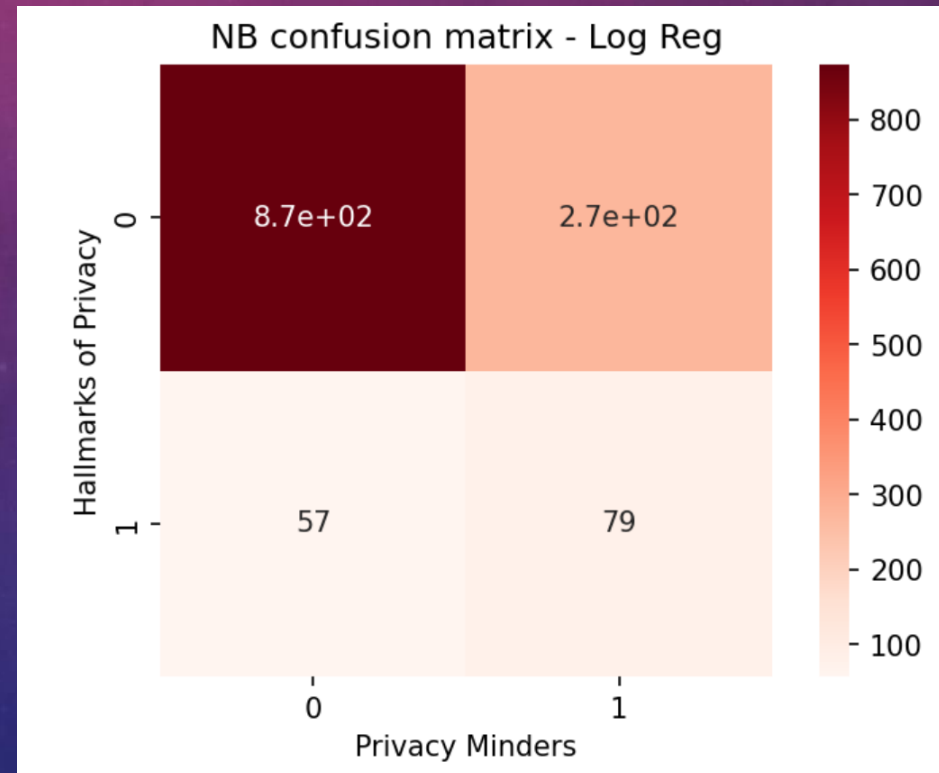
METHODOLOGY - 2

- CONVERTED ALL DATA TO BINARY TYPE
- SELECTED BERNOULLI NB, LOGISTIC REGRESSION & RANDOM FOREST CLASSIFIERS FOR MODELING
- COMPARE F1, RECALL & PRECISION SCORES BETWEEN MODELS

ANALYSIS - 1

- MODEL OF CHOICE IS LOGISTIC REGRESSION
- FEATURE IMPORTANCE EXAMINATION TO IDENTIFY HALLMARKS
- FOCUS ON FEATURES w/POSITIVE LARGE COEFFICIENTS

	Perf. Improvement	Accuracy	F1	Recall	Precision
Bernoulli NB	None	0.819	0.344	0.448	0.280
Logistic Reg.	Threshold Adjustment	0.869	0.324	0.581	0.224
Random Forest	Change Target Class Weight	0.855	0.315	0.309	0.321



ANALYSIS - 2

- 355 BINARIZED FEATURES IN TOTAL

- FEATURES w/SCORES > 2:

BENEFITCO_W49_99

TRACKCO1a_W49_99

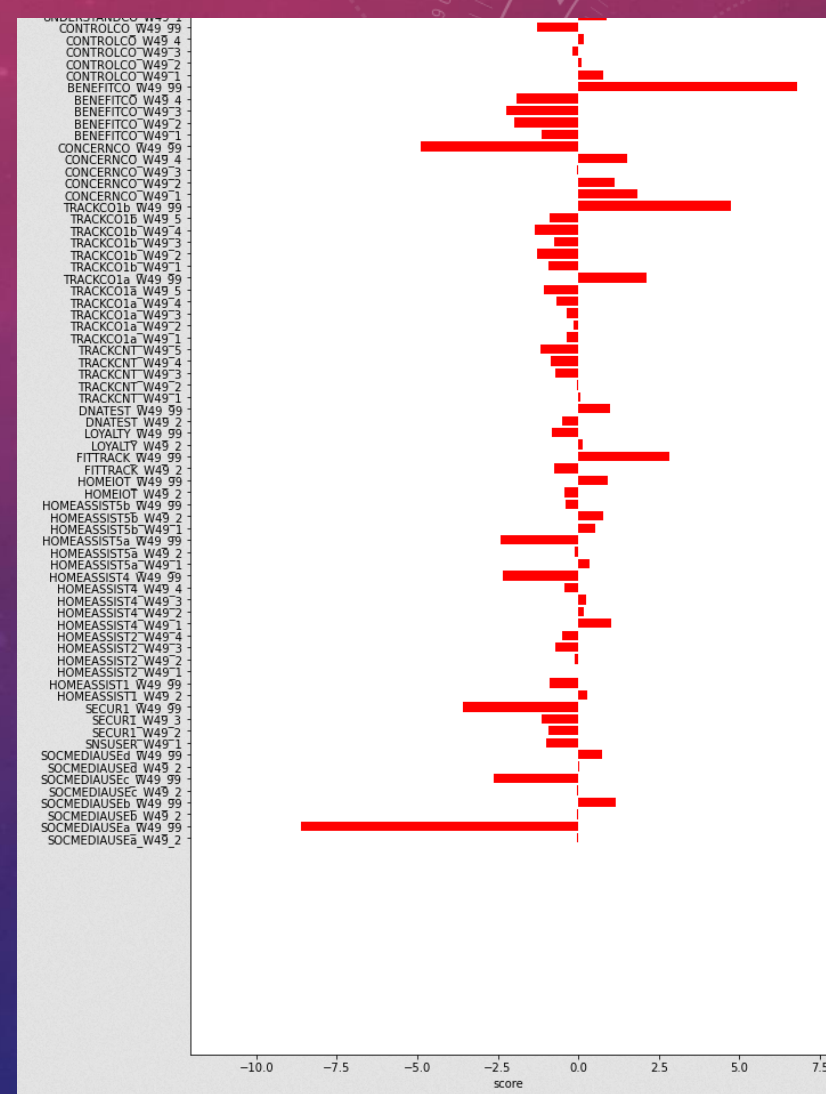
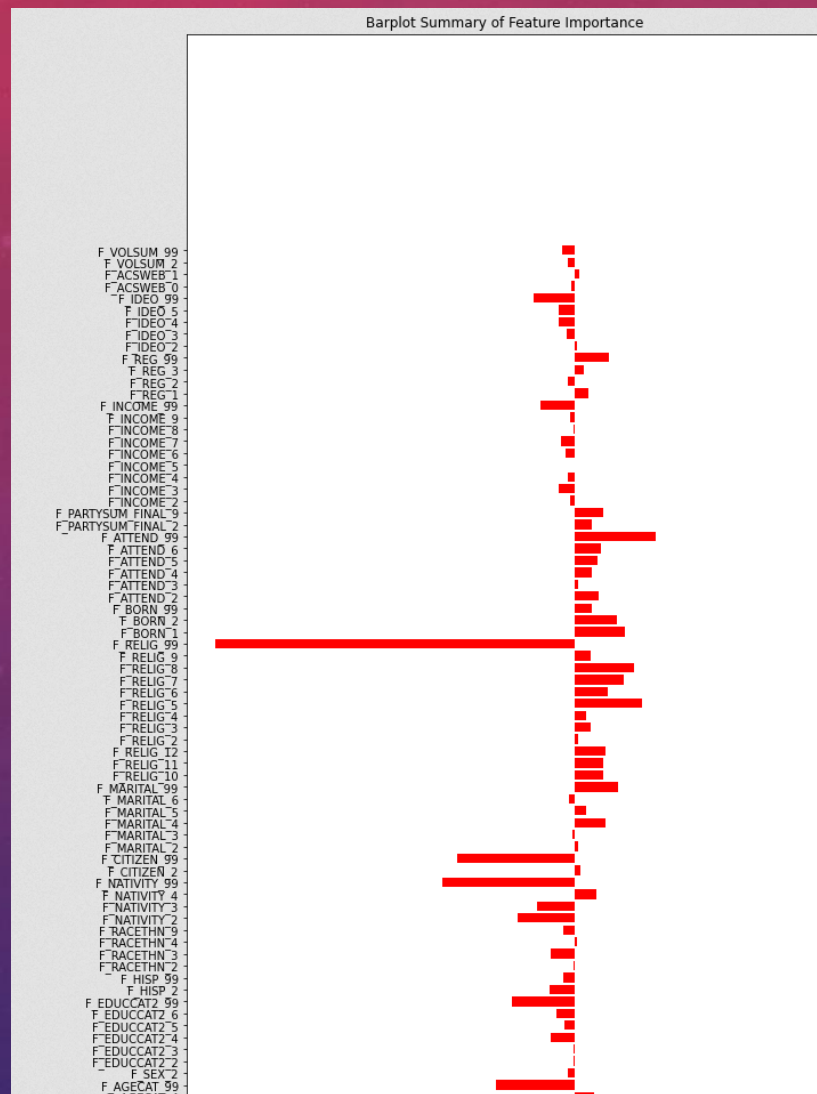
TRACKCO1b_W49_99

FITTRACK_W49_99

SOCMEDIAUSEb_W49_99

F_RELIG_5

PROFILE3_W49_3



PARTIAL FEATURE IMPORTANCE CHART

CONCLUSIONS

- FEATURES POSITIVELY DRIVING THE MODEL TEND TO BE RESPONSES WITH _99 SUFFIX (“NO ANSWER”)
- INDIVIDUALS WHO MIND THEIR ONLINE PRIVACY ARE ALSO MINDFUL ABOUT REVEALING TENDENCIES

FUTURE IMPROVEMENTS

- LOGISTIC REGRESSION & RANDOM FOREST MODELS MIGHT BENEFIT FROM OVER-SAMPLING THE TRAINING DATA
- DURING MODELING DID NOT CONSIDER USING RESPONDENT WEIGHTS DESIGNED BY PEW, CURIOUS TO SEE HOW, IF ANY, THEY MIGHT IMPACT THE SCORES

APPENDIX

- <https://www.pewresearch.org/wp-content/uploads/2018/05/Codebook-and-instructions-for-working-with-ATP-data.pdf>
- <https://www.pewresearch.org/internet/2019/11/15/data-privacy-methodology/>
- <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>
- <https://www.pewresearch.org/internet/2019/10/09/americans-and-digital-knowledge/>
- <https://www.pewresearch.org/internet/2019/09/05/more-than-half-of-u-s-adults-trust-law-enforcement-to-use-facial-recognition-responsibly/>
- <https://www.law.georgetown.edu/privacy-technology-center/publications/american-dragnet-data-driven-deportation-in-the-21st-century/>
- <https://datagy.io/sklearn-random-forests/>
- <https://medium.com/@curryrowan/simplified-logistic-regression-classification-with-categorical-variables-in-python-1ce50c4b137>
- <https://medium.com/analytics-vidhya/evaluating-a-random-forest-model-9d165595ad56>
- <https://towardsdatascience.com/methods-for-dealing-with-imbalanced-data-5b761be45a18>



THANK YOU!

DYEDBIRD@PROTONMAIL.COM