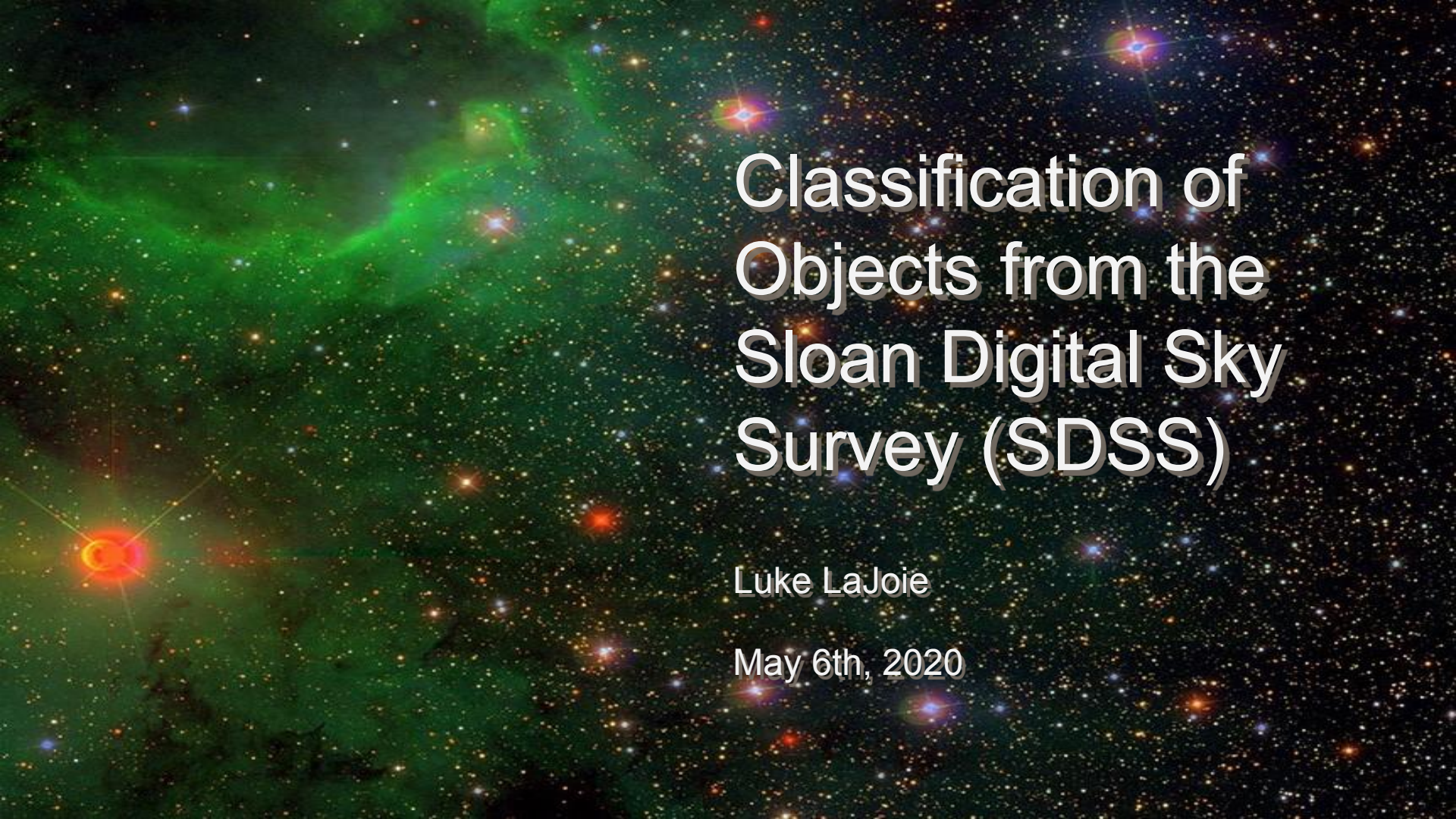


# Classification of Objects from the Sloan Digital Sky Survey (SDSS)

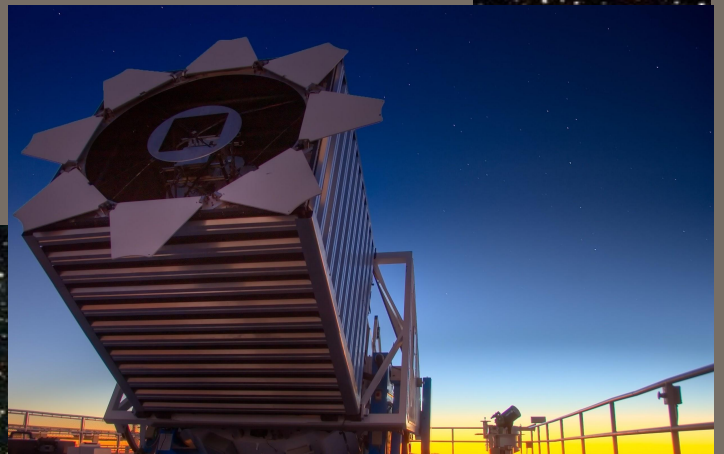
Luke LaJoie

May 6th, 2020

# Introduction

**Sloan Digital Sky Survey (SDSS)**

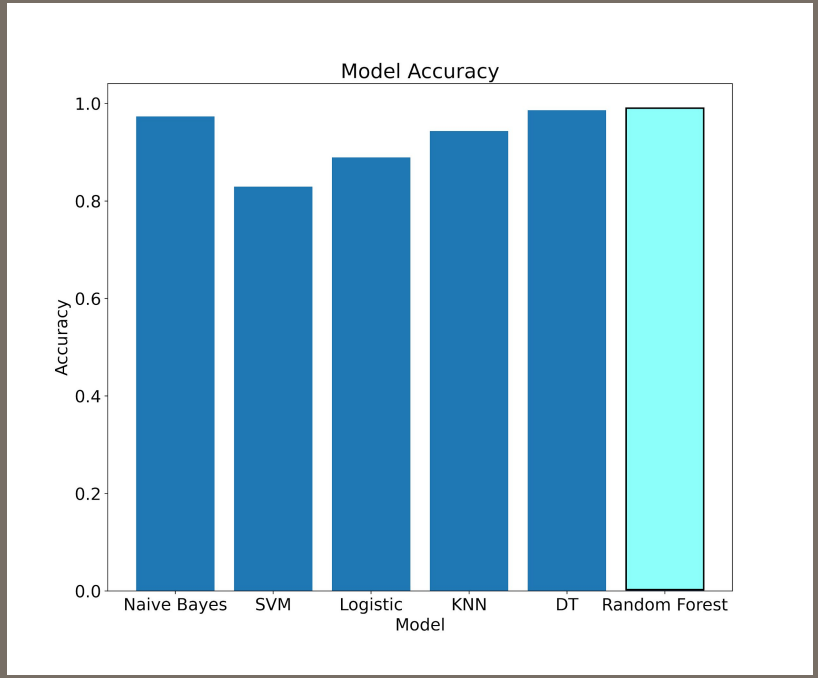
- **Detailed maps of the universe**
- **Classify: stars, galaxies, quasars (QSO)**
- **5 bands of light**
- **Redshift**
  - Related to distance



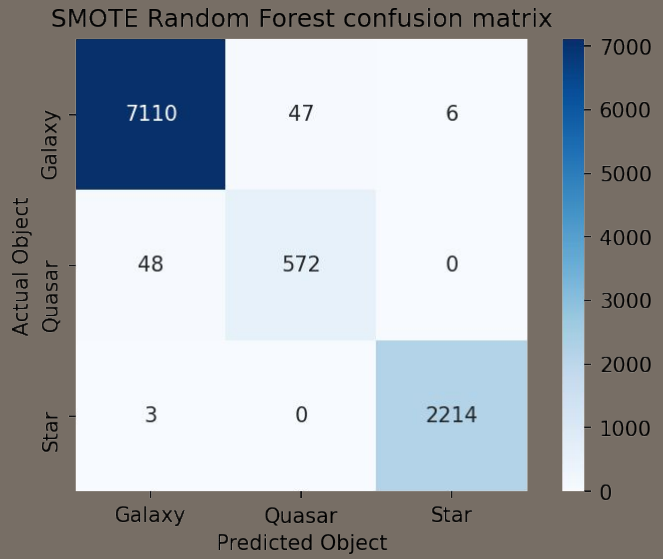
# Methodology

1. Queried SDSS online DB
2. EDA and feature engineering
3. Metric: accuracy
4. Compared models: CV, 3 classes, SMOTE

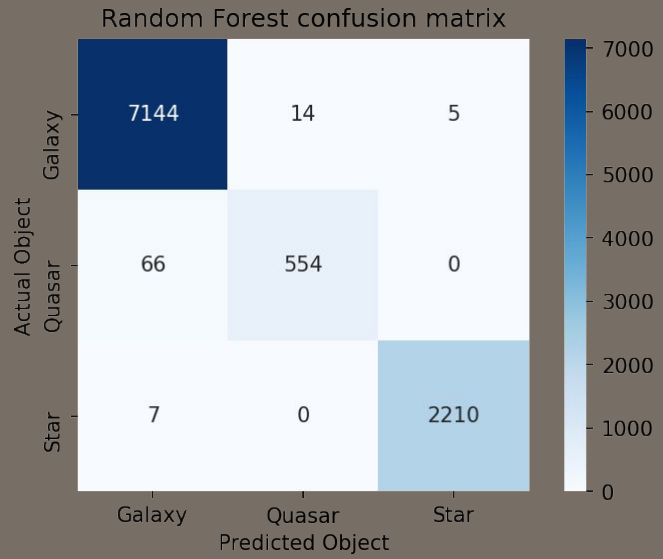
# Methodology



# Results

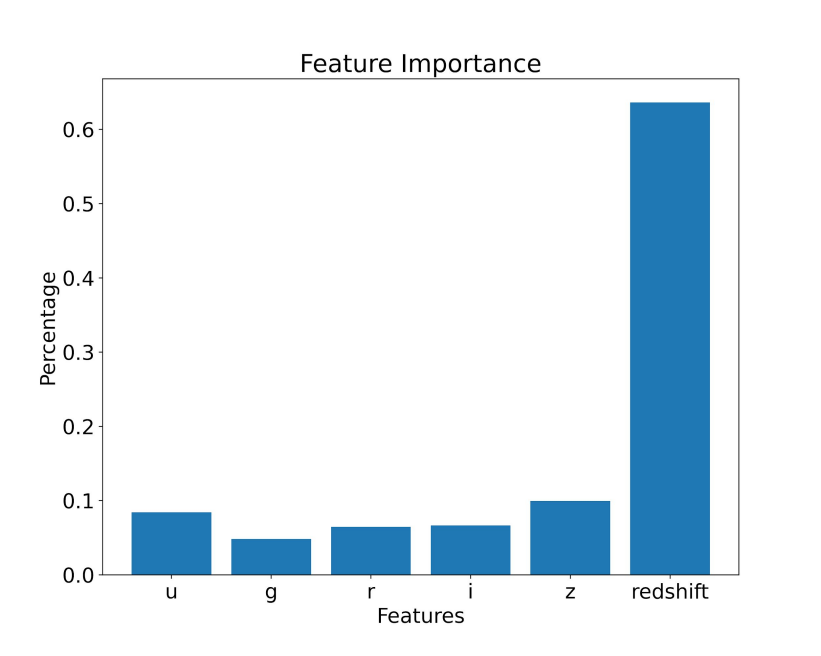



Accuracy: 0.98875




Accuracy: 0.99038

# Results


Feature Importance
Percentage
0.0
0.1
0.2
0.3
0.4
0.5
0.6
u
g
r
i
z
redshift
Features

# Results

# Conclusion

- Random Forest was the best model tested
- Redshift was the most important feature
- SMOTE increased accuracy of quasars but decreased accuracy of galaxies

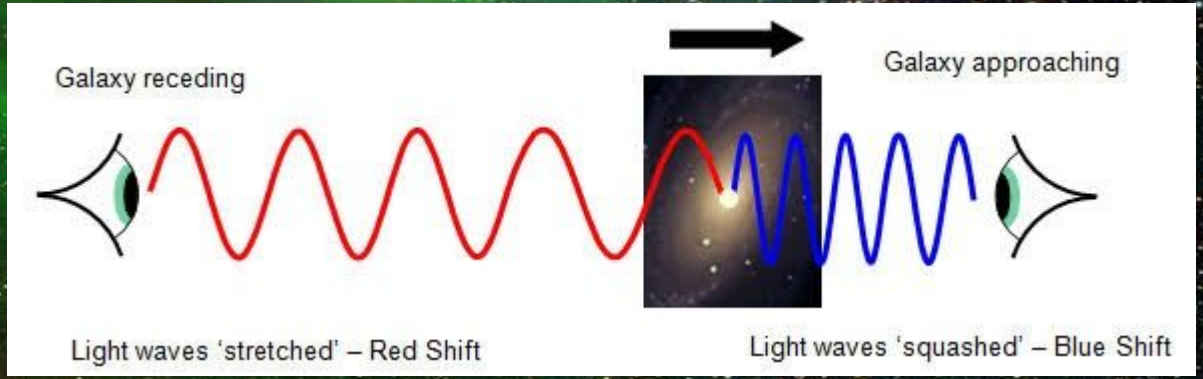
# Future Work

- **Find more accurate measure of distance**
- **Learn more about misclassified objects**
  - E.g. local group of galaxies
- **Get more features from other tables/views**

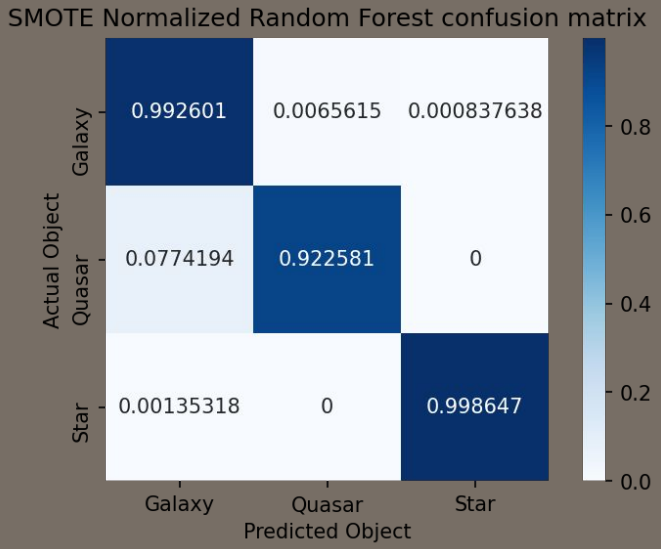

Questions?

# Appendix

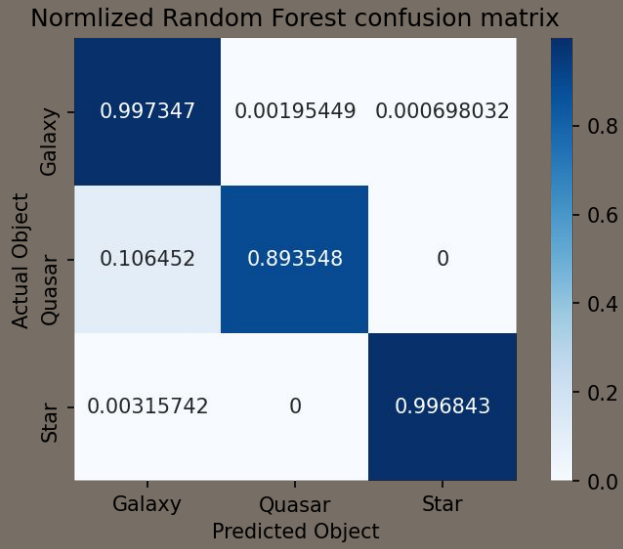
# Redshift


Galaxy receding
Galaxy approaching
Light waves 'stretched' – Red Shift
Light waves 'squashed' – Blue Shift

# Results: Normalized

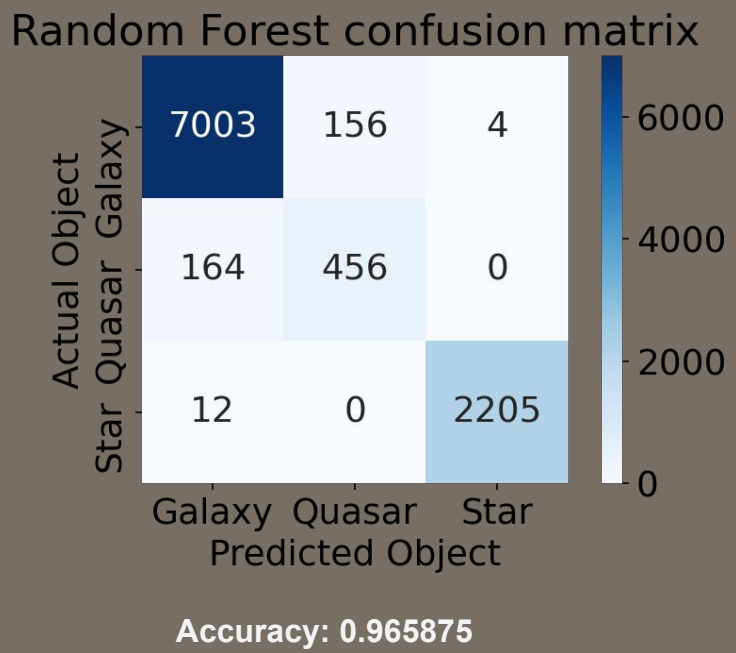



Accuracy: 0.98875



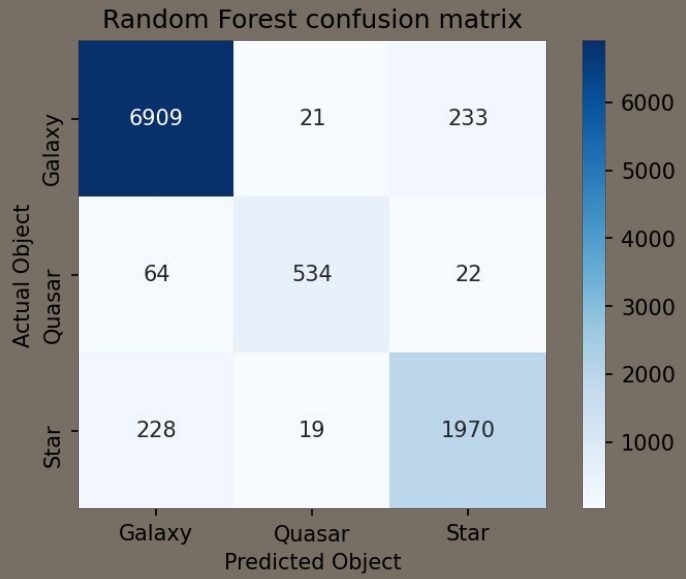

Accuracy: 0.99038

# Results: Redshift Only




**Accuracy: 0.965875**

# Results: Bands Only




**Accuracy: 0.942125**

# THANKS!

Do you have any questions?

youremail@freepik.com
+91 620 421 838
yourcompany.com