Automatic Classification of Kepler Planetary Transit Candidates Using Multilayred Neural Networks

Prabath Peiris

Capstone Project

Machine Learning Nanodegree

Udacity

Contents

1 Introduction **2**

1.1 Search for Earth-Like Planets 3

1.2 Kepler Mission 7

1.3 Photometry 8

1.4 The Transit Method of Detecting Extrasolar Planets 8

1.5 Kepler Field of View (FOV) 11

1.6 Kepler Pipeline 12

1.7 KOI and TCE Attributes 14

1.8 Missing Attributes 15

1.9 TCE Classification Labels 15

1.10 The Problem statement 16

1.11 Solution Statement 17

2 Analysis **18**

2.1 Threshold Crossing Event Catalog 18

2.2 TCE Attributes 19

2.3 Transit Fit Parameters 20

2.3.1 *tce_ror*: Ratio between planet radius and stellar radios . . . 21

2.3.2 *tce_duration*: Transit Duration 21

2.3.3 *tce_impact*: Impact Parameter 21

2.3.4 *tce_model_snr*: Transit SNR 21

2.3.5 Stellar Parameters 22

2.3.6 *tce_steff*: Stellar Effective Temperature (K) 23

2.3.7 *tce_slogg*: Stellar Surface Gravity 24

3 Methodology **27**

4 Results **28**

5 Conclusion **29**

Chapter 1

Introduction

✓ Introduction

Kepler is a space observatory launched by NASA in 2009 to discover Eath-sized exoplanets orbiting other stars [7]. Kepler mission developed several decades to answer the centuries-old question: How frequent are other Earth-like planets in Milkyway galaxy? In particular, what is the frequency of Earth-size planets in the Habitable-Zone of solar-like stars? There are three different types of exoplanets are common in our universe: gas giants, hot-super-Earths in short period orbits, and ice giants. The challenge is to find the terrestrial planets that are in the habitable zone of their host stars where liquid water might exist on the surface of the planet.

The scientific objective of the Kepler mission is to explore the structure and diversity of planetary systems. This mission surveys a large sample of stars to determine the percentage of terrestrial and large planets that are in or near the habitable zone of a wide variety of stars and determine the distribution of size and shapes of the orbits of these planets. Kepler mission also estimates how many planets are in multiple-star systems. After collecting a large number of data points, using many techniques scientists determine the properties of those stars that harbor planarity systems including the planets itself.

After preprocessing raw data, the goal is to classify each detection into one of the three different categories: Planetary Candidate (PC), Astrophysical False Positive (AFP) and None transiting phenomena (NTP). Historically this process is done by researchers looking at each observation. This is a very slow and time consuming process. During this project, I am attempting to automate this classification process using machine learning algorithms.

1.1 Search for Earth-Like Planets

✓ What Problem we are solving - Project Origin, Background problem domain

✓ - define the problem : explain in detail about the exoplanets

One of the most fundamental and intriguing philosophical questions has remained for at least 2300 years is whether life exists outside of our Solar System. With the science and technology has grown exponentially within the last couple of hundred years, it is now, we have the theoretical knowledge, practical and feasibility to seek answers to this question from a scientific perspective. NASA's Navigator Program is the long term project that over-seeing the missions related to detecting and characterization of Earth-like planets. These missions consist of multidisciplinary suite of research efforts, centering on finding exoplanets that could harbor biological activity similar to terrestrial life.

In the process of searching for Earth-like planets, we will encounter a spectrum of planets. The planetary population conceivably outnumbers stellar population since a star could host many planets such as our solar system. Unlike stars[1], plan-

[1] Vogt-Russell theorem states that the properties of a star are fully determined by the mass and chemical composision.

ets can have many different characteristics, careful examination of the planets in our solar system show us that full understanding of planets, in general, requires a working knowledge of diverse fields including star formarion, orbital mechanics, geology, grophysics, climatology, agronomy, chemistry, biology and various engineering disciplines.

we need to address a couple of questions in the search for habitable worlds:

- How do planetary systems form and evolve?
- Are there other planetary systems like our own?
- Is there life elsewhere in the Universe?

Search for habitable planets begins with an understanding of the formation of planetary systems in protoplanetary disks to learn how planetary systems form, around which types of stars planets form, how often planets form, and how the disk properties effect the distribution of final planets sizes and orbits[18].The primary mission for observing the formation of planetary systems will be the James Webb Space Telescope (JWST) [5]. JWST will allow us to study the earliest moments of star and planet formation.

In the quest of finding an extraterrestrial life form, the current observations suggest that Earth-sized rocky planets may be common; however, their abundance is quite uncertain. Kepler is a one the space-based mission[7] that is under NASA's navigator program which observes planetary systems in our solar neighborhood. Kepler also finds correlations between the presence of Eath-like planets and both stellar characteristics and the presence and orbits of giant planets. It is important to understand that no single instrument or technique is capable of finding all planetary system components around stars of all ages. Many space and ground bases missions will use complementary instrumentations and techniques to explore majority of planetary discoveries. There are four fundamental techniques will be used to determine the architecture of planetary systems: (i) *Radial Velocity Measurements* (ii) *Transit Observations*, (iii) *Astrometry* (iv) *Direct Detection.*

To answer the question "Is there life elsewhere in the universe?", we need to look for habitable planets. For us to determine whether a planet is habitable, we must build observations capable of directly detecting the light from the planet, with the planet illuminated by the light from its parent star. Direct detection of these planets is an enormous technical challenge. Another way to infer habitability of a planet is to check if the planet is located in the *Habitable Zoned* of the host star. Habitable Zone is the range in the distance from a star where liquid water could exist on the surface of a planet orbiting a star that possibly supports life. Liquid water is essential to all life on Earth, and so the definition of a habitable zone is based on the hypothesis that extraterrestrial life would share this requirement [9]. This is a very traditional definition, as a planet surface temperature may depend on other factors such as greenhouse gas abundance, its reflectivity, atmospheric and oceanic circulation, radioactive decay, and tidal heating within the planet. These energy sources can be easily allowed the planet to have subsurface liquid water reservoirs. Jupiter's moon Europa has liquid water ocean tens of kilometers below its surface that may well be habitable for some organisms. More than 20 planets, including the nearest extrasolar planet, Proxima Centauri b [1], have been found that are both roughly Earth-sized and orbiting within a Habitable Zoned of their stars.

NASA's navigator program is a scientific program whose primary goal are to detect and characterize Earth-like exoplanets and understand the formation, and

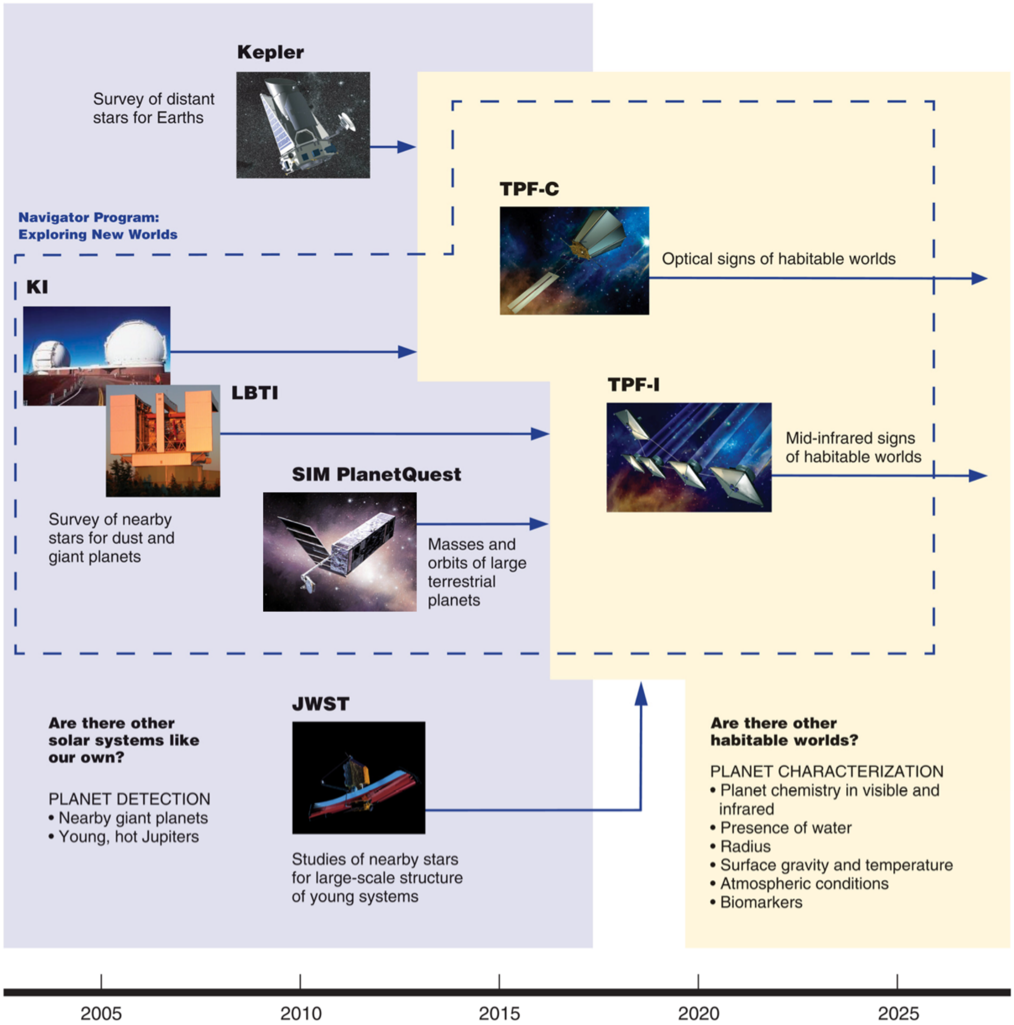


Figure 1.1: Navigator Program. National timeline including the flow of science between missions, including the *Kepler* Discovery mission and JWST. *Image Credit*: NASA

distribution of planetary systems in our Galaxy. Key features of the navigator program includes, integration of space and ground activities into a cohesive effort to find and characterize the planetary systems in our solar neighborhood. Multi-project approach to managing risk across the Navigator program by identifying the scientific and technological dependecies across the program and developing alternatives and descope to provide robustness and flexibility. The relationship between the different missions is illustrated in figure 1.1

1.2 Kepler Mission

Kepler mission and the spacecraft was named after one of history's revolutionary German astronomer and mathematician, Johannes Kepler[17]. Kepler space observatory was launched in March of 2009 into an Earth-trailing heliocentric orbit from Cape Canaveral Air Forced Station, Florida using Delta II rocket. Spacecraft design to observe fixed field of view for an extended period. Original mission duration was planned for 3.5 years; however, the mission elapsed 7 years and 11 months. On June 19, 2009, the spacecraft sent its first science data to Earth. Spacecraft continually collect science data downlink back to Earth per month. Each dataset roughly 12 gigabytes in size. On July 14, 2012, one reaction wheel (out of four) use for pointing of the spacecraft failed; however, the spacecraft only require three wheels to operate accurately. Spacecraft continually collected data from the original field of view until a second reaction wheel failed on May 11, 2013. This ends the Kepler's primary mission. At this point spacecraft no longer point to its original field of view, and this led to the "K2" follow-on mission [6] observing different fields near the elliptic orbit.

1.3 Photometry

The most basic information we can measure about celestial objects is the amount of energy that coming from it in the form of electromagnetic radiation. This quantity is known as flux. The science of measuring the flux from a celestial object is called Photometry. The photometer is an instrument that use to measure light intensity coming from an object. A Charge Coupled Devices (CCD) camera is essentially a grid of photometers that record and measure photons are coming from the sources that are in its field of view. The primary instrument of the Kepler spacecraft is a CCD photometer[12]. This CCD has 0.95-meter aperture and a 105 square degree field of view (FOV). This instrument has the sensitivity to detect Earth-like planet transit that host by a solar-like star in 6.5 hours of integration.

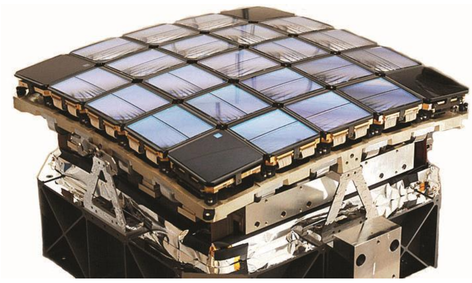


Figure 1.2: The focal plane consists of an array of 42 CCDs. Each CCD is 2.8 by 3.0 cm with 1024 by 1100 pixels. The entire focal plane contains 95 mega pixels. mage Credits: NASA Ames and Ball Aerospace

1.4 The Transit Method of Detecting Extrasolar Planets

The Kepler spacecraft detects exoplanets using transit photometry [3]. If a planet orbiting a star on the plain of view and when the it moves between the detector and the star, the light that is coming from the star partially get blocked. This event is called a "*transit*". For example, we can observe an occasional Venus or Mercury transit from Earth as a small black dot creeping across the Sun. During a transit, the flux we receive from the star reduce due to the transiting planet. When this happen, we say the planet transit the star, and can be detected using transit photometry. During these events, Kepler CCD collects raw data form of a sequence of stellar images, which are processed into "light-curves" tracking the brightness of a star over time. Light-curves are graphs that show the intensity of the light that observes from the star on the y-axis and the observation time on the x-axis (Figure 1.3). These transit data are rich with information about the planet-stellar system. The depth of the dip in the light curve and the size of the star together can be use to measure the size or the radius of the planet. The orbital period of the planet can be determined by measuring the time different between transits. Once the orbital period of the planets is known, Kepler's Third Low of planetary motion can be use to determine the average distance of the planet from its hosting star.

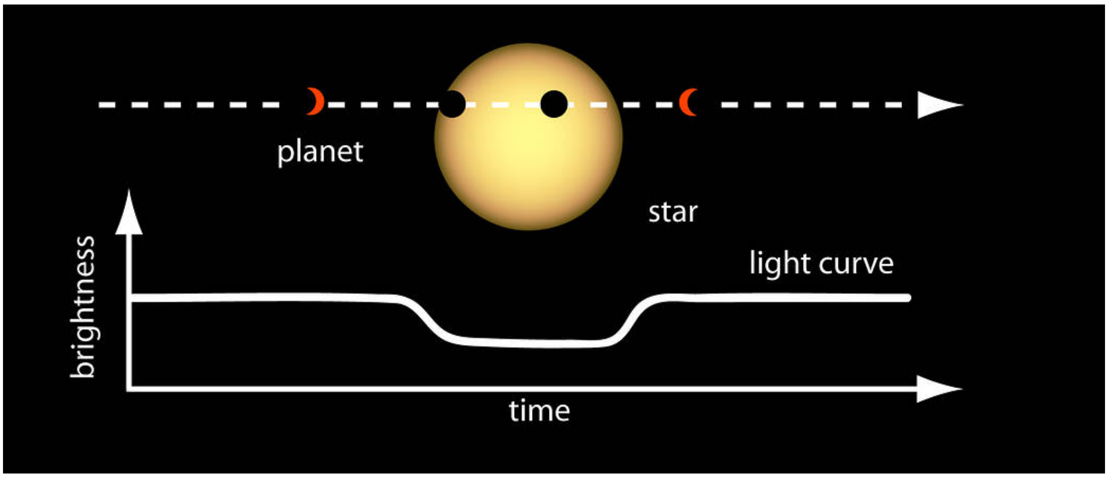


Figure 1.3: Light Curve of a Planet Transiting Its Star. Image Credit: NASA

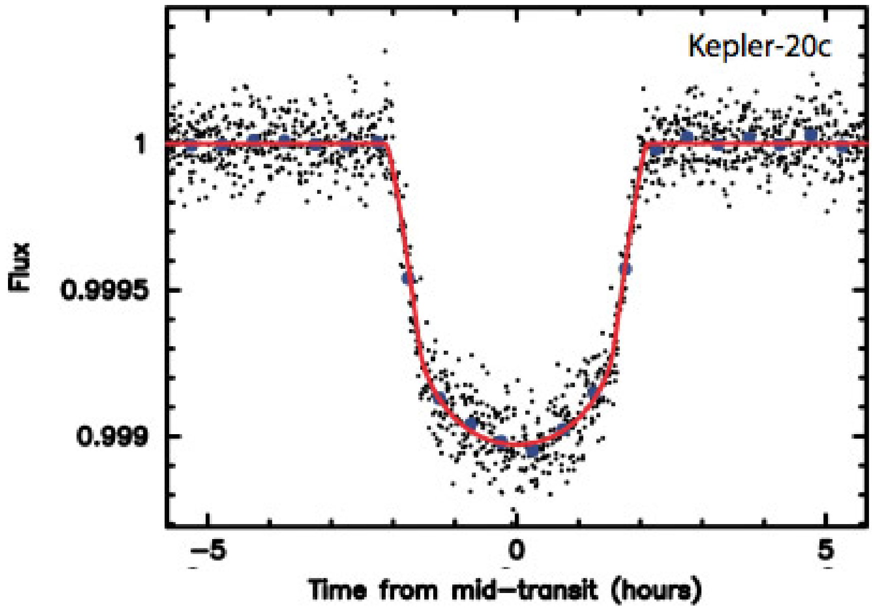


Figure 1.4: The light curve shown here was made from brightness data gathered by the Kepler Mission for discovery of planet Kepler-20c. Image Credit: NASA

Figure 1.4 shows the light curve made from the brightness data from Kepler Mission for the discovery of a planet named *Kepler-20c*. Data clearly show the flux of the host star drop notability when the planet is in transit. This is the primary signature we are looking in the transit photometry to discover planets that are in orbits around stars. These planets create this photometric signature periodically while they are orbiting around the star. To make things more complicated, the light curves of some other objects also create similar photometric signatures: eclipsing binaries stars, certain other variable stars. During the classification stage, we need to classify these object as false positives.

1.5 Kepler Field of View (FOV)

When it comes to where to look for exoplanets, there is a vast area in the sky we could point the spacecraft; however, there are a couple of technical requirements need to satisfy for the spacecraft to collect accurate data. The primary requirement is the field of view is always clear of the Sun and the Moon. This is preventing the Sunlight enter int to the CCD array. Kepler points away from the ecliptic, the line in the sky where the Sun, Moon, and the solar system planets traverse. Additionally, Kepler chose to look at an arm of the Milky Way galaxy that has stars similar in age and composition to our Sun and have the largest possible number of stars. The field of view region the Kepler mission is in the constellations Cygnus and Lyra, north of the visible band of the Milky Way. Figure 1.5 show the Field of View superimposed over Milky Way Galaxy, Figure 1.6 show the FOV relative to our Sun, Figure 1.7 show the FOV relative to Cygnus.

Figure 1.5: The Kepler field of view superimposed over a photograph of the Milky Way Galaxy

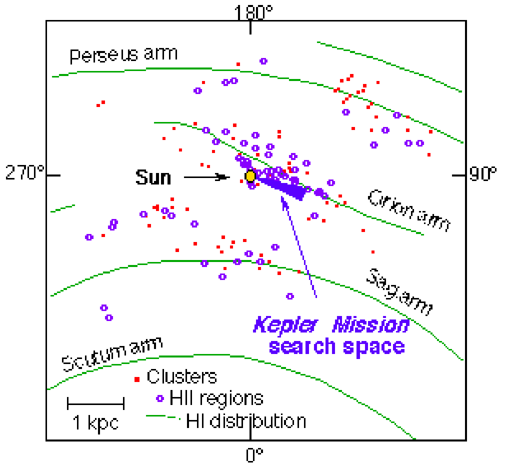


Figure 1.6: Kepler FOV in extended solar neighborhood

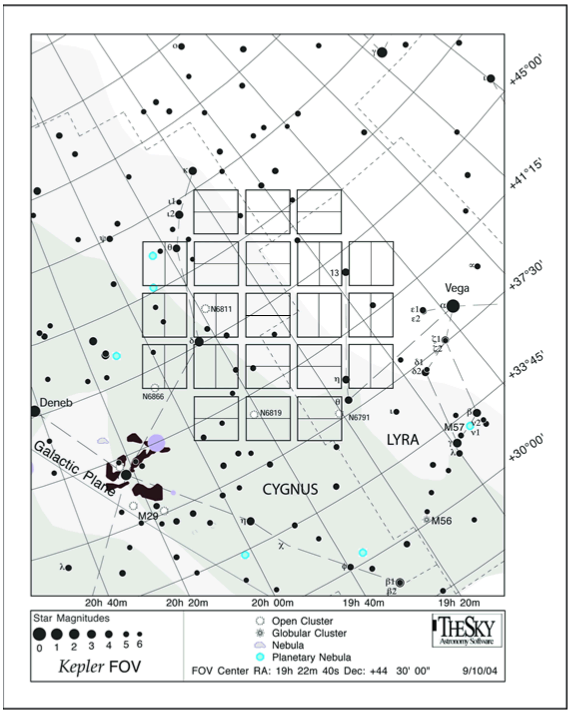


Figure 1.7: Squares are 21 CCD modules.

1.6 Kepler Pipeline

The Kepler Pipeline [7] is a data reduction pipeline used for translating the Kepler raw pixel data into possible transiting planet detections. Kepler mission perform photometric observations of carefully selected stars (around 156,000) using its 115 deg^2 field of view (FOV) as reviewed in Borucki et el. (2010) [2] and Koch et al. (2010) [8]. The Kepler Mission Science Operations Center (SOC) at NASA Ames Research Center performs major functions on these datasets including calibrate CCD array, download data (light curves) from the spacecraft periodically, remove systematic noise [15] and perform statistical tests to reject false positives and establish accurate statistical confidence in each detection.

[4].

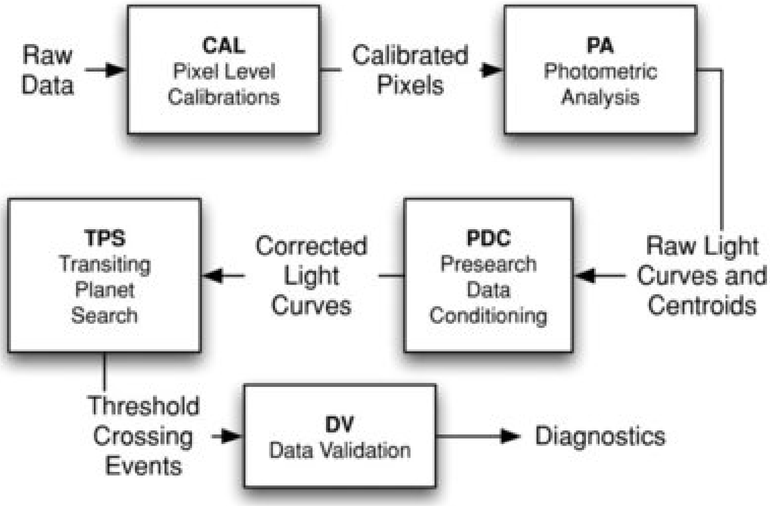


Figure 1.8: Data flow diagram for the SOC Science Pipeline. Image Credit: The American Astronomical Society

Figure 1.8 show the major steps and modules of the pipeline. In particular the last two modules of the pipeline; those that identify as Threshold Crossing Events (TCEs) and their subsequent transit model fiting. TCE is a sequence of significant, periodic, planet transit-like features in the light curve of a target star. Transiting Planet Search module takes systematic error-correlated light curve for a star and seach parameter space for possible transit signatures. This module outputs a TCE or say that does not exists TCE event on the target star. This produce smaller subset of target stars what is given to the Data Validation (DV) module. DV module takes initial TCE and gaps the transit signatures from the light curve and uses the Transiting Planet Search to find additional TCEs on the same target star. This process repeats until it finds all the TCEs on given star. More details on this process explained by Mandel and Agol (2002) [10] and Claret and Bloemen (2011) [4].

TPS algorithm detects transit-like features in light curves by applying noise compensating, wavelet-based matched filtering. TPS characterized the power spectral density (PSD) of the observation noise as a function of time to implement a whitening filter in the wavelet domain. The trial transit pulse is whitened and correlated against the whitened flux time series. Features with correlations above the threshold of 7.1σ are flagged as potential threshold crossing events and subjected to additional tests in TPS to guard against false alarms.

Algorithm searches a parameter space with varying transit durations D and produce a Single Event Statistics (SES) time series that is the significance of the detection of the reference transmit pulse centered at that particular time for each D

$$SES(t) = N(t)/\sqrt{D(t)} \tag{1.1}$$

$\sqrt{D(t)}$, is the expected signal to noise ratio of a signal that exactly matches the template pulse and $N(t)$ is the correlated time series.

Multiple Event statistics (MES) is constructed that characterizes a significant detection in a search over varying orbital period p and epochs (phase) t_0 by folding $N(n)$ and $D(n)$. MES $> 7.1\sigma$ may produce a TCE if it also passes additional statistical tests. SES and MES are the basis of some of the attributes used in the training set.

1.7 KOI and TCE Attributes

The Threshold Crossing Events (TCE) catalog contain a sequence of transit-like features in the flux time series of a given target star. These TCE data can download from NASA Exoplanet Archive databases [2], and also the detail description of the table fields [3] are listed as public data. Kepler Object of Interest (KOI) catalog contains object data including many attributes. The detail attributes are listed on NASA Exoplanet Archive website [4], and dataset can be download from the arcive tables [5]. All these data cab be download as bulk using data tools available in the archive website.

1.8 Missing Attributes

In some circumstances, there are missing attributes in the dataset due to missing information in the stellar catalogs; the data validation fit fails to converge, or a processing timeout is reached. These missing attributes can be fill using reasonable methods such as mean, median, and in case of missing stellar attributes can be filled using sun's parameters. These various substitutions will be tested when training the network.

1.9 TCE Classification Labels

Each Threshold Crossing Event (TCE) is subject to a vetting process performed by the Kepler TCE Review Team (TCERT). During the triage (Initial) vetting stage, all TCEs are partition into two different sets: Problematic Ligh Curves that has instrumental noise and Kepler Object of Interest. KOI is a TCE that contains convincing transit-like features that do not present obvious evidence that the TCE

[2] `http://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=q1_q17_dr24_tce`

[3] `http://exoplanetarchive.ipac.caltech.edu/docs/API_tce_columns.html`

[4] `http://exoplanetarchive.ipac.caltech.edu/docs/API_kepcandidate_columns.html`

[5] `http://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=cumulative`

was generated from non-transiting phenomena such as instrumental noise. These KOIs moves to next level of the vetting process performed by individuals manually inspecting light curves using detection statistics. Any indication they see the signal came from an eclipsing binary star or more complex forms of instrumental noise removed from the list. TCEs that survive this removal process are classified as Planet Candidates (PC).

We can identify three different types of classification labels in the processed dataset: Planetary Candidates (PC), Astrophysical False Positives (AFP) and non-transiting phenomena (NTP). PCs are confirmed as planets, statistically validated as planets or determined to be a planet candidate by the TCERT. AFPs are those TCEs that have been shown to be eclipsing binary stars or have shown evidence that the transiting object being detected is not located around the target tart. NTP are those TCEs that failed the initial vetting process.

1.10 The Problem statement

Kepler is a single instrument spacecraft that collect most contiguous and long-running photometric time series possible. Kepler observe approximately 170,000 stars simultaneously while it is operating. The fundamental objective of the Kepler mission is to detect a large number of transiting exoplanets. The ultimate goal of the primary mission was a characterize the frequency of exoplanets on diameter, orbital period and host star. The Manual classification of the findings of Kepler object has proven very time-consuming. The new space-based, transit photometry missions such as K2 [6], TESS [14], and PLATO 2.0 [13] also produce a large number of the dataset that demands some level of automation to do the classification.

Using machine learning classification techniques, we can speed up the process and provide a more continuous rating of planarity candidates. There are various of machine learning classification techniques has been applied to Kepler dataset including random forests, SVM, K-mean clustering [16, 11]. In this project, we are attempting to train a Multilayered Neural Network to identify the potential planetary candidates in the Kepler dataset.

1.11 Solution Statement

Machine learning techniques contribute a way to automate some step of exoplanet discovery. The TCE vetting process is a tedious and time-consuming (mostly a manual) process. Modern astronomical observational instruments generate a large amount of data within a short period of observational time. These observations may contain such a crucial events that need future follow-up observations using other telescopes (using other wavelengths); hence, processing these time series data and extracting meaningful information is a time sensitive process. Solution to this is to process Kepler data using machine learning algorithms to express the classification while reducing human errors. I am attempting to trained Neural Network to process the TCE catalog to automate the classification process.

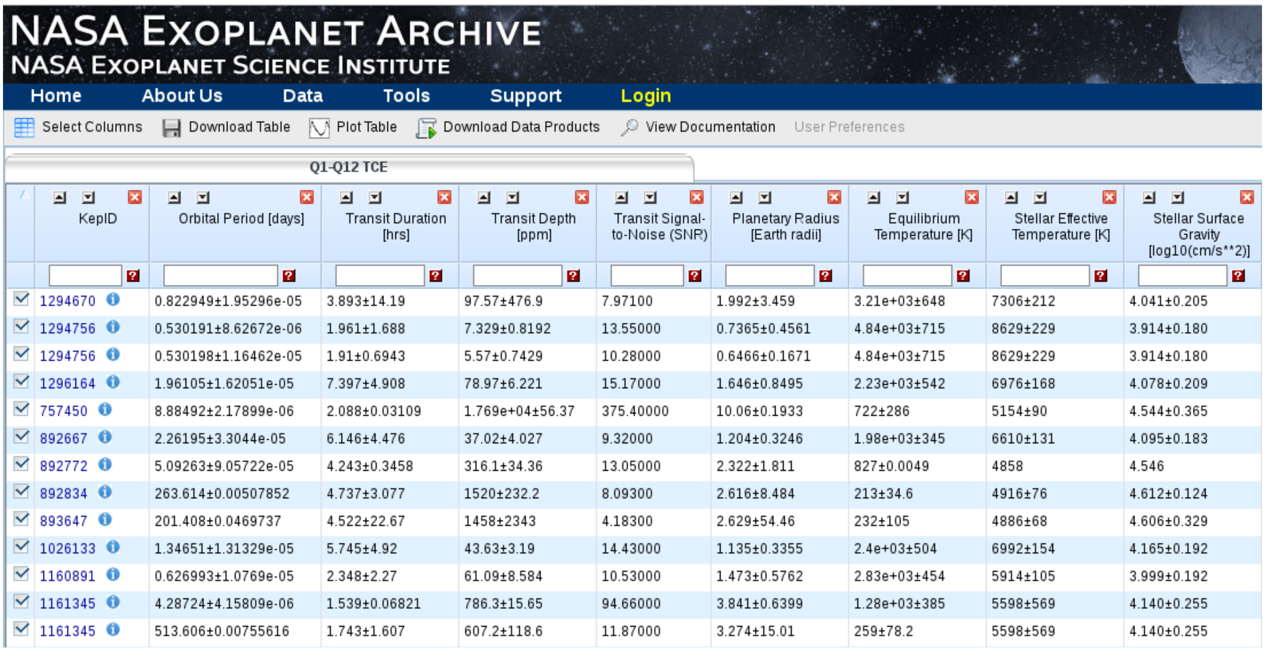
Chapter 2

Analysis

A Threshold-Crossing Events (TCE) are built using flux time series that has a sequence of transit-like features. The flux time series of a target star that resembles the signature of transiting planet with a high degree of confidence passed on for further analysis. Each TCE identified by a Kepler ID (KID). In the case of a star that holds multiple planets may have a many TCEs.

2.1 Threshold Crossing Event Catalog

Figure 2.1 show a sample of TCE datalog from Exoplanet Archive from NASA Exoplanet Science Institute. This is an interactive table that has exoplanet archive including information provided by the original sources. Each row that has unique KeplerID is a TCE. In Figure 2.1 show only a few columns out of many. Each column is an observed or a computed data point that related to the given TCE. Using the upper left hand "Select Column" menu, we can select any fields to appear on the table. By hovering over a KeplerID, you can explore more data that is belong to this particular TCE (show in Figure 2.2). Kepler target overview page show more complete stellar parameters including stellar 2MASS images. Kepler TCE



KepID	Orbital Period [days]	Transit Duration [hrs]	Transit Depth [ppm]	Transit Signal-to-Noise (SNR)	Planetary Radius [Earth radii]	Equilibrium Temperature [K]	Stellar Effective Temperature [K]	Stellar Surface Gravity [log10(cm/s**2)]
1294670	0.822949±1.95296e-05	3.893±14.19	97.57±476.9	7.97100	1.992±3.459	3.21e+03±648	7306±212	4.041±0.205
1294756	0.530191±8.62672e-06	1.961±1.688	7.329±0.8192	13.55000	0.7365±0.4561	4.84e+03±715	8629±229	3.914±0.180
1294756	0.530198±1.16462e-05	1.91±0.6943	5.57±0.7429	10.28000	0.6466±0.1671	4.84e+03±715	8629±229	3.914±0.180
1296164	1.96105±1.62051e-05	7.397±4.908	78.97±6.221	15.17000	1.646±0.8495	2.23e+03±542	6976±168	4.078±0.209
757450	8.88492±2.17899e-06	2.088±0.03109	1.769e+04±56.37	375.40000	10.06±0.1933	722±286	5154±90	4.544±0.365
892667	2.26195±3.3044e-05	6.146±4.476	37.02±4.027	9.32000	1.204±0.3246	1.98e+03±345	6610±131	4.095±0.183
892772	5.09263±9.05722e-05	4.243±0.3458	316.1±34.36	13.05000	2.322±1.811	827±0.0049	4858	4.546
892834	263.614±0.00507852	4.737±3.077	1520±232.2	8.09300	2.616±8.484	213±34.6	4916±76	4.612±0.124
893647	201.408±0.0469737	4.522±22.67	1458±2343	4.18300	2.629±54.46	232±105	4886±68	4.606±0.329
1026133	1.34651±1.31329e-05	5.745±4.92	43.63±3.19	14.43000	1.135±0.3355	2.4e+03±504	6992±154	4.165±0.192
1160891	0.626993±1.0769e-05	2.348±2.27	61.09±8.584	10.53000	1.473±0.5762	2.83e+03±454	5914±105	3.999±0.192
1161345	4.28724±4.15809e-06	1.539±0.06821	786.3±15.65	94.66000	3.841±0.6399	1.28e+03±385	5598±569	4.140±0.255
1161345	513.606±0.00755616	1.743±1.607	607.2±118.6	11.87000	3.274±15.01	259±78.2	5598±569	4.140±0.255

Figure 2.1: Threshold Crossing Event Catalog: `http://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=q1_q12_tce`

Overview page shows in-depth details of the TCE, this includes Kepler data validation reports and Kepler Planet Detection Metrics and time series data including various downloading options. We can download all the time series data from these catalogs.

2.2 TCE Attributes

Initial TCE attribute set contains 237 attributes that are based on the wavelet matched filter use by TPS, transit model fitting, difference image centroids, and some additional tests. Each of these attributes has different strength of prediction values. The importance of these attributes is selected using historical literature on

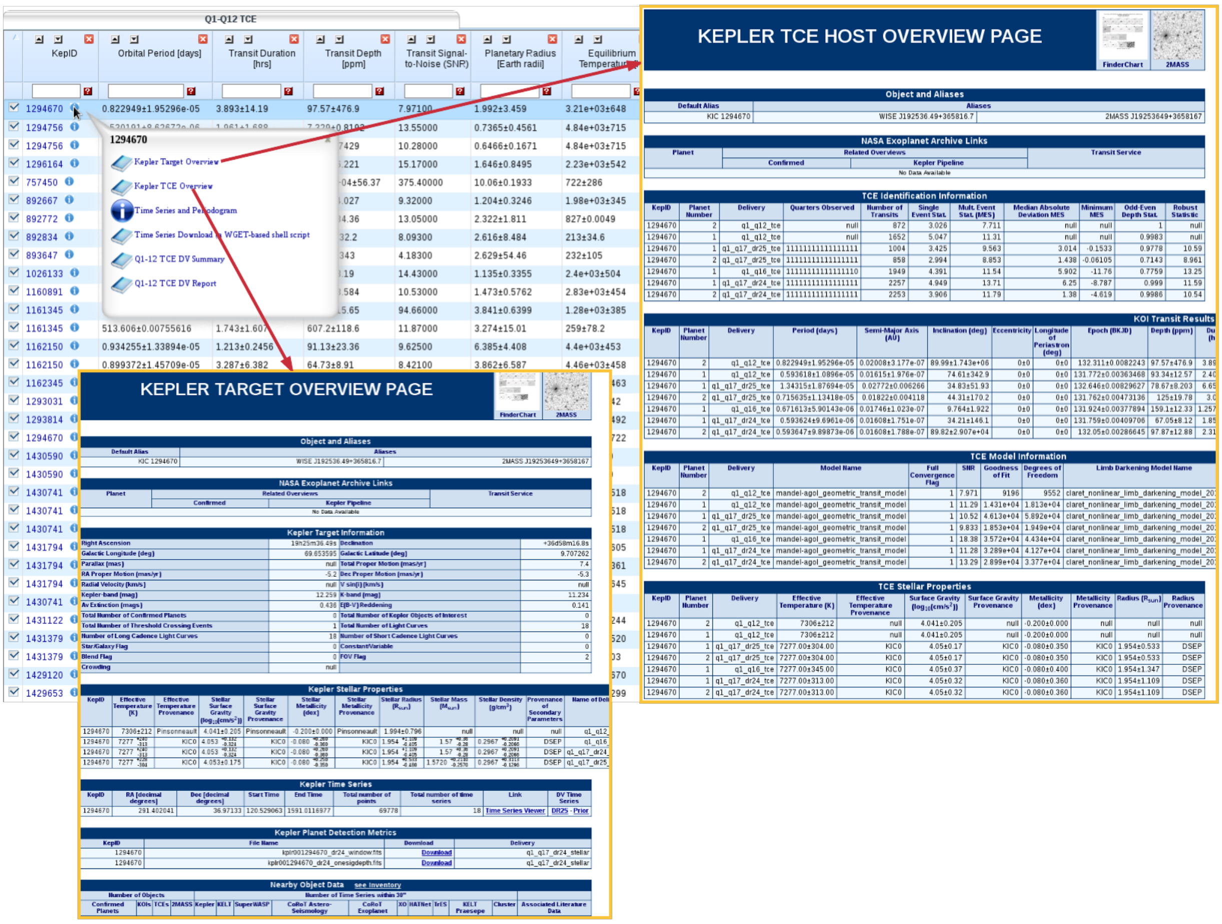


Figure 2.2: Threshold Crossing Event Catalog Details

attributes and performing a Principal Component Analysis. We discuss some of the most important attributes in this section.

2.3 Transit Fit Parameters

This paper presents exact analytic formulate for the eclipse of a star described by quadratic or nonlinear limb darkening. The Kepler Project derives transit parameters from best-fit parameters produced by this analytic formula. Some of the transit parameters are calculated directly; others are derived from the best-fit parameters. Limb-darkening coefficients are fixed and pre-calculated from host stellar properties.

2.3.1 *tce_ror*: Ratio between planet radius and stellar radios

tce_ror attribute calculated by planet radius divided by its hosting the stellar radius. Both planet and stellar radius are in Eath-radii units. Plot 2.3 show the histogram of the *tce_ror* attribute in the TCE catalog (x-axis is in log scale and y-axis is in linear)

2.3.2 *tce_duration*: Transit Duration

tce_duration: The duration of the observed transits. Duration is measured from the first contact between the planet and star until the last contact. The duration is measured by hours. The plot 2.4 show the distribution of the transit duration of the TCE catalog. We can observe the most of the transit are fall under 10-hour duration.

2.3.3 *tce_impact*: Impact Parameter

tce_impact: The sky-projected distance between the center of the stellar disc and the center of the planet disc at conjunction, normalized by the stellar radius.

2.3.4 *tce_model_snr*: Transit SNR

Transit depth normalized by the mean uncertainty in the flux during the transits. Plot 2.5 show the distribution of the SNR.

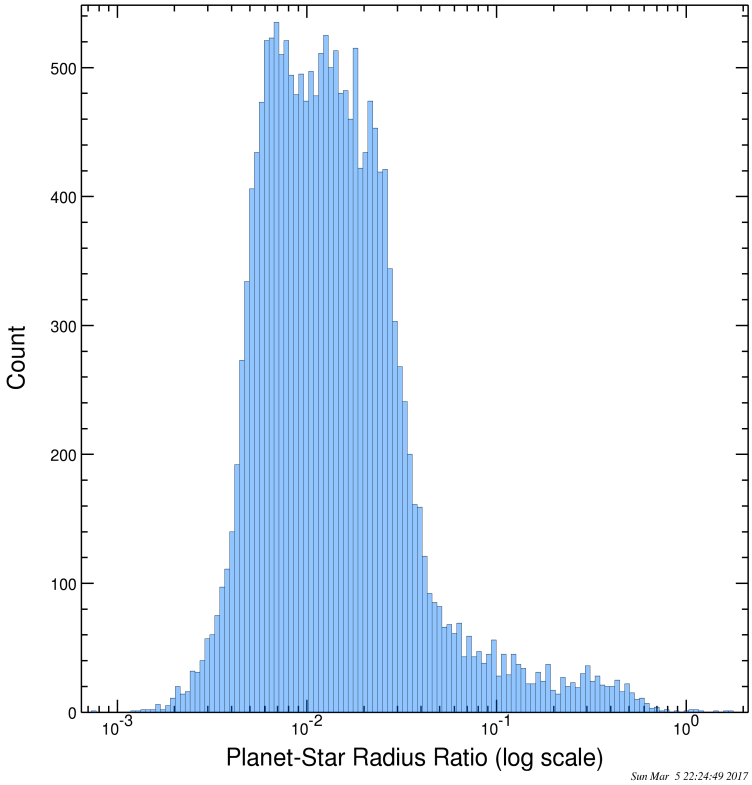


Figure 2.3: Ratio between planet radius and stellar radios

2.3.5 Stellar Parameters

Best-fit transit parameters are normalized to the size of the host star. The size of the hosting star is determined by its radius. Physical planet parameters may be derived by scaling to the star's size and temperature. Stellar effective temperature, surface gravity, metallicity, radius, mass, and age should comprise a consistent set. This section describe and visualize some of the stellar parameters from the TCE catalog.

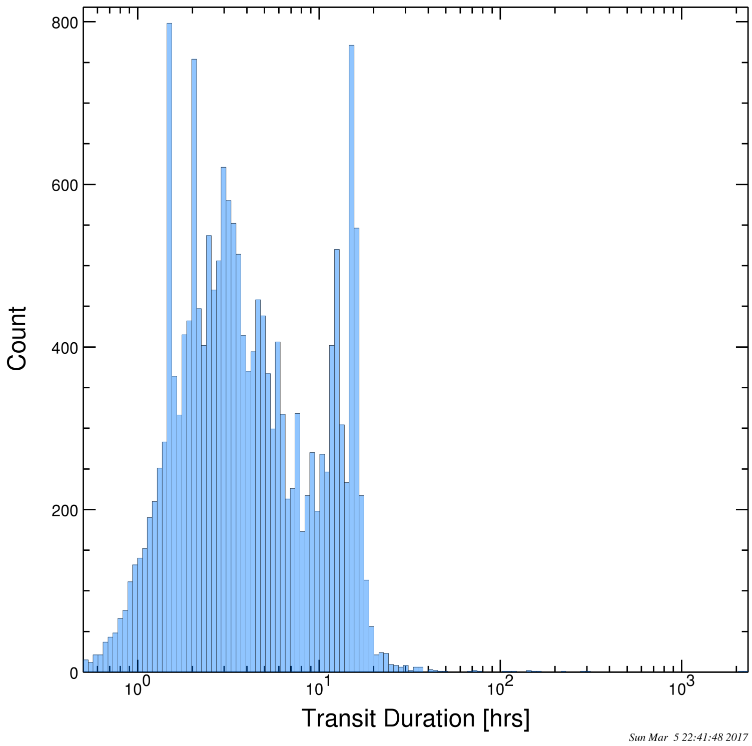


Figure 2.4: Transit Duration

2.3.6 *tce_steff*: Stellar Effective Temperature (K)

Plot 2.6 show the photospheric temperature of the stars in the TCE catalog. Kepler mission target sun-like stars and the photospheric temperature of the sun is 5,775 K. If we plot the sun on the 2.6, it will fall close to the apparently large number count that shows on the plot. This 2.6 show the TCE catalog stellar temperatures are falling between 3000K and 17000K

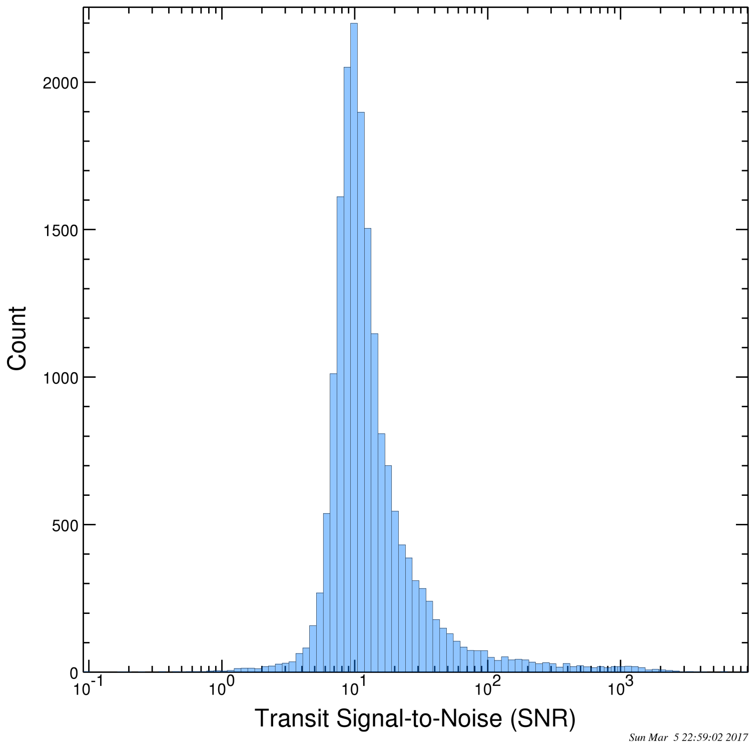


Figure 2.5: Transit SNR

2.3.7 *tce_slogg*: Stellar Surface Gravity

The base-10 logarithm of the acceleration due to gravity at the surface of the star. The surface gravity of our sun is 274.0 cm/s^2 (4.44 log10(cm/s^2)). Plot 2.7 show the distribution of the TCE catalog stellar surface gravity.

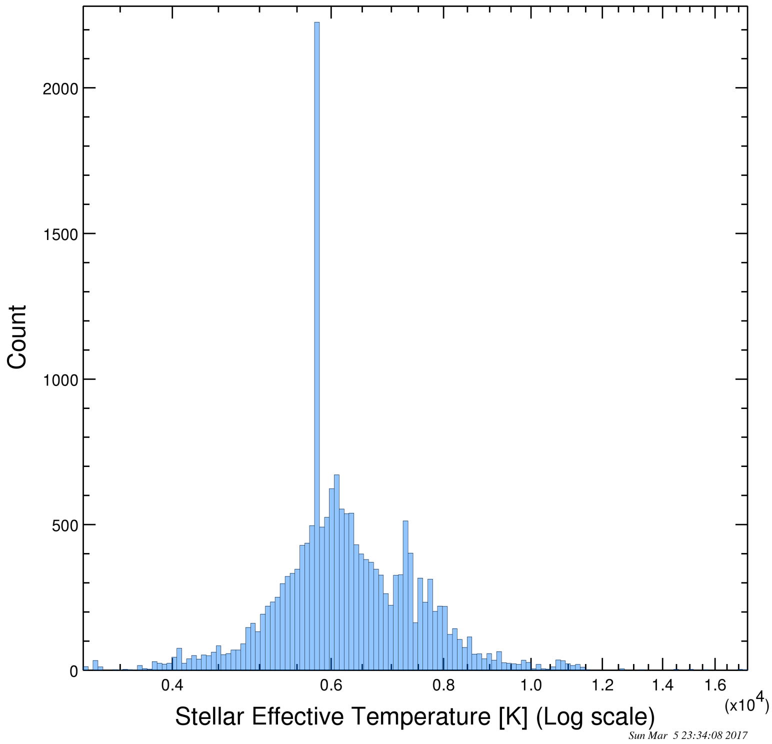


Figure 2.6: Stellar Effective Temperature (K)

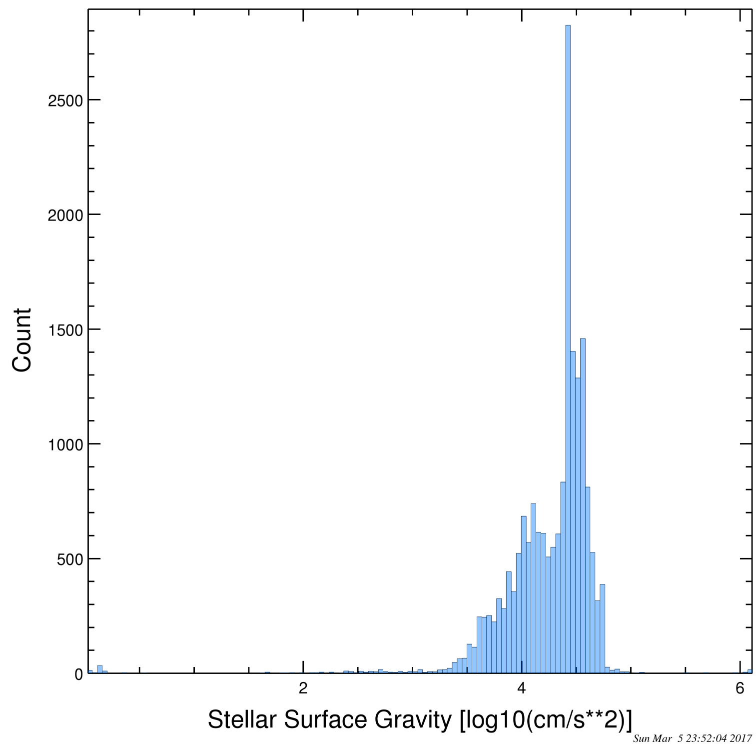


Figure 2.7: Stellar Surface Gravity log10(cm/s^2)

Chapter 3

Methodology

★ Data Preprocessing

★ - talk about all the data preprocessings steps. PCA, and collect the the top 10 PCs and how that is define, show the table.

★ Implementation

★ - defien the trainng and testing data

★ - Explain the cording process and any complications

★ Refinement

★ - Explain how the process improved. show the initial with someother activation function with bad outputs and bad loss funcitons. show some bad initial results.

Chapter 4

Results

★ Model evaluation and validation

★ - Explain the final model paramters, explain

★ - Explain the and show the loss funstion, show the confision matrix

★ justification

★ - Compare the final results to the bench mark - compare the confustion matrix to the once we have on the paper, also compare to the graph the score value, explain and show the problem is adiculately solved

Chapter 5

Conclusion

★ Conclusion discussion - based on the conclusion section of the paper

★ Describe the entier process - highlight whats interesting and whats difficult, interesting - Astrophysics can replace the physical data processing piplelines with machine learning and allow real time dataprocessing with the large data comming from the telescopes

★ How can I Improve - this is where explain we can now do this using light curves instead using TCE catalog

Bibliography

[1] G. Anglada-Escudé, P. J. Amado, J. Barnes, Z. M. Berdiñas, R. P. Butler, G. A. L. Coleman, I. de La Cueva, S. Dreizler, M. Endl, B. Giesers, S. V. Jeffers, J. S. Jenkins, H. R. A. Jones, M. Kiraga, M. Kürster, M. J. López-González, C. J. Marvin, N. Morales, J. Morin, R. P. Nelson, J. L. Ortiz, A. Ofir, S.-J. Paardekooper, A. Reiners, E. Rodríguez, C. Rodríguez-López, L. F. Sarmiento, J. P. Strachan, Y. Tsapras, M. Tuomi, and M. Zechmeister. A terrestrial planet candidate in a temperate orbit around Proxima Centauri. *nat*, 536:437–440, August 2016.

[2] William J. Borucki, David Koch, Gibor Basri, Natalie Batalha, Timothy Brown, Douglas Caldwell, John Caldwell, Jørgen Christensen-Dalsgaard, William D. Cochran, Edna DeVore, Edward W. Dunham, Andrea K. Dupree, Thomas N. Gautier, John C. Geary, Ronald Gilliland, Alan Gould, Steve B. Howell, Jon M. Jenkins, Yoji Kondo, David W. Latham, Geoffrey W. Marcy, Søren Meibom, Hans Kjeldsen, Jack J. Lissauer, David G. Monet, David Morrison, Dimitar Sasselov, Jill Tarter, Alan Boss, Don Brownlee, Toby Owen, Derek Buzasi, David Charbonneau, Laurance Doyle, Jonathan Fortney, Eric B. Ford, Matthew J. Holman, Sara Seager, Jason H. Steffen, William F. Welsh, Jason Rowe, Howard Anderson, Lars Buchhave, David Ciardi, Lucianne Walkowicz,

William Sherry, Elliott Horch, Howard Isaacson, Mark E. Everett, Debra Fischer, Guillermo Torres, John Asher Johnson, Michael Endl, Phillip MacQueen, Stephen T. Bryson, Jessie Dotson, Michael Haas, Jeffrey Kolodziejczak, Jeffrey Van Cleve, Hema Chandrasekaran, Joseph D. Twicken, Elisa V. Quintana, Bruce D. Clarke, Christopher Allen, Jie Li, Haley Wu, Peter Tenenbaum, Ekaterina Verner, Frederick Bruhweiler, Jason Barnes, and Andrej Prsa. Kepler planet-detection mission: Introduction and first results. *Science*, 327(5968):977–980, 2010.

[3] D. Charbonneau, T. M. Brown, D. W. Latham, and M. Mayor. Detection of Planetary Transits Across a Sun-like Star. *apjl*, 529:L45–L48, January 2000.

[4] A. Claret and S. Bloemen. VizieR Online Data Catalog: Limb-darkening coefficients (Claret+, 2011). *VizieR Online Data Catalog*, 352, March 2011.

[5] Jonathan P. Gardner, John C. Mather, Mark Clampin, Rene Doyon, Matthew A. Greenhouse, Heidi B. Hammel, John B. Hutchings, Peter Jakobsen, Simon J. Lilly, Knox S. Long, Jonathan I. Lunine, Mark J. Mccaughrean, Matt Mountain, John Nella, George H. Rieke, Marcia J. Rieke, Hans-Walter Rix, Eric P. Smith, George Sonneborn, Massimo Stiavelli, H. S. Stockman, Rogier A. Windhorst, and Gillian S. Wright. The james webb space telescope. *Space Science Reviews*, 123(4):485–606, 2006.

[6] S. B. Howell, C. Sobeck, M. Haas, M. Still, T. Barclay, F. Mullally, J. Troeltzsch, S. Aigrain, S. T. Bryson, D. Caldwell, W. J. Chaplin, W. D. Cochran, D. Huber, G. W. Marcy, A. Miglio, J. R. Najita, M. Smith, J. D.

Twicken, and J. J. Fortney. The K2 Mission: Characterization and Early Results. *pasp*, 126:398–408, April 2014.

[7] J. M. Jenkins, D. A. Caldwell, H. Chandrasekaran, J. D. Twicken, S. T. Bryson, E. V. Quintana, B. D. Clarke, J. Li, C. Allen, P. Tenenbaum, H. Wu, T. C. Klaus, C. K. Middour, M. T. Cote, S. McCauliff, F. R. Girouard, J. P. Gunter, B. Wohler, J. Sommers, J. R. Hall, A. K. Uddin, M. S. Wu, P. A. Bhavsar, J. Van Cleve, D. L. Pletcher, J. A. Dotson, M. R. Haas, R. L. Gilliland, D. G. Koch, and W. J. Borucki. Overview of the Kepler Science Processing Pipeline. *apjl*, 713:L87–L91, April 2010.

[8] D. G. Koch and et al. Kepler Mission Design, Realized Photometric Performance, and Early Science. *apjl*, 713:L79–L86, April 2010.

[9] R. K. Kopparapu, R. M. Ramirez, J. SchottelKotte, J. F. Kasting, S. Domagal-Goldman, and V. Eymet. Habitable Zones around Main-sequence Stars: Dependence on Planetary Mass. *apjl*, 787:L29, June 2014.

[10] K. Mandel and E. Agol. Analytic Light Curves for Planetary Transit Searches. *apjl*, 580:L171–L175, December 2002.

[11] S. D. McCauliff, J. M. Jenkins, J. Catanzarite, C. J. Burke, J. L. Coughlin, J. D. Twicken, P. Tenenbaum, S. Seader, J. Li, and M. Cote. Automatic Classification of Kepler Planetary Transit Candidates. *apj*, 806:6, June 2015.

[12] E. A. Petigura and G. W. Marcy. Identification and Removal of Noise Modes in Kepler Photometry. *pasp*, 124:1073, October 2012.

[13] H. Rauer, C. Catala, C. Aerts, T. Appourchaux, W. Benz, A. Brandeker, J. Christensen-Dalsgaard, M. Deleuil, L. Gizon, M.-J. Goupil, M. Güdel, E. Janot-Pacheco, M. Mas-Hesse, I. Pagano, G. Piotto, D. Pollacco, Ċ. Santos, A. Smith, J.-C. Suárez, R. Szabó, S. Udry, V. Adibekyan, Y. Alibert, J.-M. Almenara, P. Amaro-Seoane, M. A.-v. Eiff, M. Asplund, E. Antonello, S. Barnes, F. Baudin, K. Belkacem, M. Bergemann, G. Bihain, A. C. Birch, X. Bonfils, I. Boisse, A. S. Bonomo, F. Borsa, I. M. Brandão, E. Brocato, S. Brun, M. Burleigh, R. Burston, J. Cabrera, S. Cassisi, W. Chaplin, S. Charpinet, C. Chiappini, R. P. Church, S. Csizmadia, M. Cunha, M. Damasso, M. B. Davies, H. J. Deeg, R. F. Díaz, S. Dreizler, C. Dreyer, P. Eggenberger, D. Ehrenreich, P. Eigmüller, A. Erikson, R. Farmer, S. Feltzing, F. de Oliveira Fialho, P. Figueira, T. Forveille, M. Fridlund, R. A. García, P. Giommi, G. Giuffrida, M. Godolt, J. Gomes da Silva, T. Granzer, J. L. Grenfell, A. Grotsch-Noels, E. Günther, C. A. Haswell, A. P. Hatzes, G. Hébrard, S. Hekker, R. Helled, K. Heng, J. M. Jenkins, A. Johansen, M. L. Khodachenko, K. G. Kislyakova, W. Kley, U. Kolb, N. Krivova, F. Kupka, H. Lammer, A. F. Lanza, Y. Lebreton, D. Magrin, P. Marcos-Arenal, P. M. Marrese, J. P. Marques, J. Martins, S. Mathis, S. Mathur, S. Messina, A. Miglio, J. Montalban, M. Montalto, M. J. P. F. G. Monteiro, H. Moradi, E. Moravveji, C. Mordasini, T. Morel, A. Mortier, V. Nascimbeni, R. P. Nelson, M. B. Nielsen, L. Noack, A. J. Norton, A. Ofir, M. Oshagh, R.-M. Ouazzani, P. Pápics, V. C. Parro, P. Petit, B. Plez, E. Poretti, A. Quirrenbach, R. Ragazzoni, G. Raimondo, M. Rainer, D. R. Reese, R. Redmer, S. Reffert, B. Rojas-Ayala, I. W. Roxburgh, S. Salmon, A. Santerne, J. Schneider, J. Schou, S. Schuh, H. Schunker, A. Silva-Valio, R. Silvotti, I. Skillen, I. Snellen, F. Sohl, S. G. Sousa, A. Sozzetti,

D. Stello, K. G. Strassmeier, M. v Svanda, G. M. Szab, A. Tkachenko, D. Valencia, V. Van Grootel, S. D. Vauclair, P. Ventura, F. W. Wagner, N. A. Walton, J. Weingrill, S. C. Werner, P. J. Wheatley, and K. Zwintz. The PLATO 2.0 mission. *Experimental Astronomy*, 38:249–330, November 2014.

[14] G. R. Ricker, J. N. Winn, R. Vanderspek, D. W. Latham, G. Á. Bakos, J. L. Bean, Z. K. Berta-Thompson, T. M. Brown, L. Buchhave, N. R. Butler, R. P. Butler, W. J. Chaplin, D. Charbonneau, J. Christensen-Dalsgaard, M. Clampin, D. Deming, J. Doty, N. De Lee, C. Dressing, E. W. Dunham, M. Endl, F. Fressin, J. Ge, T. Henning, M. J. Holman, A. W. Howard, S. Ida, J. Jenkins, G. Jernigan, J. A. Johnson, L. Kaltenegger, N. Kawai, H. Kjeldsen, G. Laughlin, A. M. Levine, D. Lin, J. J. Lissauer, P. MacQueen, G. Marcy, P. R. McCullough, T. D. Morton, N. Narita, M. Paegert, E. Palle, F. Pepe, J. Pepper, A. Quirrenbach, S. A. Rinehart, D. Sasselov, B. Sato, S. Seager, A. Sozzetti, K. G. Stassun, P. Sullivan, A. Szentgyorgyi, G. Torres, S. Udry, and J. Villasenor. Transiting Exoplanet Survey Satellite (TESS). In *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, volume 9143 of *procspie*, page 914320, August 2014.

[15] J. C. Smith, M. C. Stumpe, J. E. Van Cleve, J. M. Jenkins, T. S. Barclay, M. N. Fanelli, F. R. Girouard, J. J. Kolodziejczak, S. D. McCauliff, R. L. Morris, and J. D. Twicken. Kepler Presearch Data Conditioning II - A Bayesian Approach to Systematic Error Correction. *pasp*, 124:1000–1014, September 2012.

[16] G. Torres, D. M. Kipping, F. Fressin, D. A. Caldwell, J. D. Twicken, S. Ballard, N. M. Batalha, S. T. Bryson, D. R. Ciardi, C. E. Henze, S. B. Howell, H. T. Isaacson, J. M. Jenkins, P. S. Muirhead, E. R. Newton, E. A. Petigura, T. Bar-

clay, W. J. Borucki, J. R. Crepp, M. E. Everett, E. P. Horch, A. W. Howard, R. Kolbl, G. W. Marcy, S. McCauliff, and E. V. Quintana. Validation of 12 Small Kepler Transiting Planets in the Habitable Zone. *apj*, 800:99, February 2015.

[17] J.R. Voelkel. *Johannes Kepler and the New Astronomy.* Oxford Portraits in Science. Oxford University Press, USA, 2001.

[18] A. N. Youdin and F. H. Shu. Planetesimal Formation by Gravitational Instability. *apj*, 580:494–505, November 2002.