Статус на 14 августа 2016 года

1. Изучил статью XHIP: An Extended Hipparcos Compilation. В ней описывается новый каталог XHIP. В его основе лежат данные HIPNEWCAT (2007), объединенные с данными из многих других источников. Если в оригинальном HIPPARCOS 39807 звезд имеют класс светимости, то в XHIP их уже 61315. Вот ссылка на статью <https://yadi.sk/d/5vHd63lmomK7f>
2. Повторил все свои вычисления по данным каталога XHIP. Принципиально ничего не изменилось, но повысилась достоверность всех расчетов.
3. Затем я занялся проверкой гипотезы, что “красные карлики имеют большие собственные движения, а красные гиганты – маленькие собственные движения”. Для этого я классифицировал звезды с B-V > 0.6 по одному фактору – собственному движению. Получилось, что если звезда гигант, то мы с большой вероятностью скажем, что она гигант. Но если она карлик, то классификатор с большой достоверностью не сможет сказать, карлик она, или гигант. То есть зависимость между собственным расстоянием и типом звезды является слабой.

Я вернулся к проблеме классификации карликов и гигантов с B-V < 0.6. Напомню, что раньше наш классификатор полагал, что все звезды с B-V < 0.6 являются карликами. Это происходило из-за того, что карликов в этой области было в 6 раз больше, чем гигантов. Я выкинул 5/6 случайных карликов из обучающего множества, т.е. сравнял по количеству карликов и гигантов в обучающем множестве. Так же, я вместо SVM взял другой алгоритм классификации – ансамбль деревьев решений. В итоге, получилось разделить карликов и гигантов с точностью/полнотой 70%, что является хорошим результатом.

1. Начал изучать работы, связанные с применением методов машинного обучения к задачам астрономии. Таких работ оказалось достаточно много, вот некоторые из них:

- **DATA MINING AND MACHINE LEARNING IN ASTRONOMY** <https://arxiv.org/pdf/0906.2173v2.pdf>

- **Astronomy with scikit-learn** <http://www.astroml.org/sklearn_tutorial/_downloads/sklearn_tutorial.pdf>

- **Advances in Machine Learning and Data Mining for Astronomy**

<http://press.princeton.edu/chapters/s10159.pdf>

Материалов на эту тему полно – надо поизучать и найти какую-нибудь интересную тему.

1. Начал изучать Python-библиотеку **astroML**, специализирующуюся на задачах машинного обучения в астрономии http://www.astroml.org. Она базируется на Python-библиотеке для машинного обучения **scikit-learn** http://scikit-learn.org. Я понял, что сейчас машинное обучение в астрономии программируют в основном именно на Python, поэтому весь новый код я начну писать именно на Python, а не на Java, как раньше. То есть доработки, связанные с обработкой данных Hipparcos, я продолжу делать на Java в старом проекте, а новые задачи на новых данных буду программировать на Python уже в новом проекте. Авторы библиотеки написали книгу по моей теме <https://www.amazon.com/Statistics-Mining-Machine-Learning-Astronomy/dp/0691151687>. Я хочу ее заказать. Есть ли смысл это делать через библиотеку мат-меха?

P.S. Кстати, в этой библиотеке есть уже встроенный HealPix <http://www.astroml.org/book_figures/chapter1/fig_healpix_ex.html>