

Механизм испарения может быть запущен разными методами. Например, он реализуется при работе магнетронных распылительных систем с жидкофазными мишенями. Вещество мишени расплавляется за счёт того, что создаётся теплоизоляция её от остальных элементов катодного узла и магнитной системы, которые должны охлаждаться в процессе работы МРС. Тепловая энергия, которая выделяется в мишени под действием плазмы, остается в ней. При определённых условиях её оказывается достаточно для расплавления вещества мишени и создания интенсивного испарения на её поверхности.

Другой возможный способ создания испарения – повышение плотности тока ионов плазмы МРС, бомбардирующей поверхность мишени и переносящих ей основную долю энергии, вкладываемой в разряд. Он может быть реализован при работе сильноточных импульсных магнетронных систем, а также при использовании ассистирующего ионного пучка, действие которого усиливает процессы ионизации в плазме магнетронного диода.

Цель исследований состояла в том, чтобы проанализировать возможности МРС разных видов (сильноточных импульсных МРС и МРС с жидкофазными мишенями) для создания испарения на поверхности мишеней. Решение этой задачи выполнено расчётным путём. Для этого разработаны математические модели тепловых и эрозионных процессов на катодном узле на основе уравнений теплопроводности с учётом фазовых переходов первого рода. Получены данные о закономерностях эрозии мишеней при возникновении фазовых превращений первого рода и о параметрах МРС, при которых обеспечиваются наилучшие показатели интенсивности и энергоэффективности получения эрозионного материала. Показано, что включение испарения приводит к увеличению как минимум на порядок коэффициента эрозии поверхности мишени и скорости роста покрытий по сравнению с режимами работы МРС, в которых имеет место только столкновительное распыление.

Выполнены экспериментальные исследования скорости роста металлический покрытий при использовании МРС с жидкофазными мишенями. Расчётные результаты и данные экспериментов хорошо согласуются, что свидетельствует о корректности сформулированных физических представлений и полученных закономерностей.

## **ИСПОЛЬЗОВАНИЕ ПРОГРАММНЫХ ИНСТРУМЕНТОВ SAS ДЛЯ ПОДГОТОВКИ БОЛЬШИХ ДАННЫХ**

Е.И. Губин

Национальный исследовательский Томский политехнический университет

E-mail: gubine@tpu.ru

## **USING SAS SOFTWARE TOOLS FOR PREPARING BIG DATA**

E. I. Gubin

National Research Tomsk Polytechnic University

***Annotation.** Formation of basic competencies in the preparation of source data in the field of large volumes of data for predictive analysis in the SAS environment.*

Для прогнозного анализа «чистота» исходных данных носит исключительно важный характер. Многие авторы отмечают критический характер этого и предлагают конкретные шаги по сбору и подготовке исходных слабо структурированных больших данных [1, 2].

Однако предложенные мероприятия носят по большей мере рекомендательный и теоретический характер, так как выявление факта «некорректности» исходных данных само по себе не простая задача, особенно касаясь реально больших данных.

В настоящей работе автор хотел бы представить современные программные инструменты (коды) SAS для проведения статистического анализа при выявлении некорректных (ошибочных) входных данных.

Важной исходной процедурой является трансфер (импорт) данных в среду SAS, где предполагается проводить основные вычисления [3,4]. Ниже приведен пример кода импорта исходных данных (data set: athlete\_events в формате csv) в библиотеку среды SAS (data set WORK.test1).

```
PROC IMPORT
DATAFILE="C:\Users\gubine\Desktop\xls\athlete_events.csv"
OUT=WORK.test1
REPLACE
DBMS=csv;
GETNAMES=YES;
RUN;
```

Если следовать логике «чистке» исходных данных, изложенной в работах [1,2], то выявление отсутствующих значений («missing») для текстовых и числовых атрибутов в среде SAS, будет следующим:

```
proc format;
value $missfmt ' '= 'Missing' other= 'Not Missing';
value missfmt . = 'Missing' other= 'Not Missing';
run;

proc freq data=test1;
format _CHAR_ $missfmt.; /* apply format for the duration of this PROC */
tables _CHAR_ / missing missprint nocum nopercnt;
format _NUMERIC_ missfmt.;
tables _NUMERIC_ / missing missprint nocum nopercnt;
run;
```

В данном случае в качестве исходного файла взят data set WORK.test1

Наличие одинаковых наблюдений влияет на коэффициенты регрессии, увеличивая дисперсию модели, поэтому дублирующие наблюдения должны быть найдены и удалены из анализа. Выявление в исходных данных дублирующих строк (наблюдений) в среде SAS, будет следующим:

```
proc sort data= test1/**not duplicate***/
nodupkey out= test1_1; by ID;
run;
```

Выбросы в исходных данных («outliers») – это аномальные значения, выделяющиеся из общей выборки. Логистическая регрессия чувствительна к выбросам, поэтому их обработка является очень важным шагом подготовки данных.

```
proc univariate data=test1 robustscale plot; /**outliers***/
var income age;
run;
```

Выявление мультиколлинеарности между входными переменными носит важный характер для прогнозных моделей (особенно регрессий) в смысле переобучения модели на обучающей выборке, что может привести к преувеличенной статистической значимости входных переменных на целевую функцию [5].

Для анализа мультиколлинеарности во входных переменных, используем специальные процедуры SAS:

```
proc princomp data=test1 /**correlation***/
```

```
outstat=test1_stat noprint;
```

```
run;
```

Для экспорта данных, полученных в среде SAS, в наиболее популярную для бизнеса Excel, код SAS будет следующим:

```
proc export data=test1
```

```
outfile="C:\Users\gubine\Desktop\xls\резерв_11.xls"
```

```
dbms=xls
```

```
replace;
```

```
run;
```

В данной работе предложена методика подготовки данных для построения прогнозных моделей классификации с использованием технологий SAS. Этапы подготовки данных включают в себя следующие шаги: **1.** проверку исходных данных на ошибки (описки), **2.** на отсутствие данных (“missing”), **3.** на выбросы данных (“outliers”), **4.** на наличие дублирующих строк (наблюдений), **5.** на проверку исходных объясняющих переменных (атрибутов) на мультиколлинеарность.

### СПИСОК ЛИТЕРАТУРЫ

1. Губин Е.И. Методика подготовки больших данных для прогнозного анализа // Наука и бизнес: пути развития. – 2020. – № 3(105). – С. 33–35.
2. Губин Е.И. Методология подготовки больших данных для прогнозного анализа// Современные технологии, экономика и образование: сборник трудов Всероссийской научно-методической конференции – Томск, 2019. – 139 с. – С. 25–28.
3. Вершинин А.С., Губин Е.И. Применение инструмента DATA MINING для оценки кредитоспособности заемщика // Информационные технологии в науке, управлении, социальной сфере и медицине: труды V Международной конференции. – Томск, 2018. – Т.2. – С. 18–21.
4. Вершинин А.С., Губин Е.И. Использование инструментов SAS для оценки рисков заемщиков // Молодежь и современные информационные технологии: Труды XVI Международной научно-практической конференции студентов, аспирантов и молодых ученых, – Томск, 2018. – С. 379–380.
5. Руководство по кредитному скорингу /под ред. Элизабет Мэйз; пер. с англ. И.М. Вороненко. – Минск: Гревцов Паблицер, 2008. – 464 с.

### РЕАЛИЗАЦИЯ СБОРА ДАННЫХ В ПРОГРАММНОМ ОБЕСПЕЧЕНИИ СТЕНДА ДЛЯ ИСПЫТАНИЯ НАНОСЕНСОРОВ

Д.И. Коваль, В.Л. Ким

Национальный исследовательский Томский политехнический университет

E-mail: deniskoval12@gmail.com

### IMPLEMENTATION OF DATA COLLECTION IN THE SOFTWARE OF THE NANOSENSOR TEST BENCH

D.I. Koval, V.L. Kim

National Research Tomsk Polytechnic University

***Annotation:** The process of implementing the software of the lower level of the stand for testing nanosensors is described, the stages of work on the project, the flowchart with the project operation cycles are given, and the modes of packet transmission are described (sending packets to the device and receiving packets by the device).*

### Ведение.

Реализация модульных приборов, для дальнейшей их интеграции целостный аппаратно-программный комплекс (АПК), который служит для выполнения задач по замеру биосигналов, в наше время имеет очень большую важность. Предпосылки в